

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:35:16 : Search time 30.13 Seconds

(without alignments)
409,200 Million cell updates/sec

Title: US-10-048-197-2

Perfect score: 558

Sequence: 1 MLNHNIRLRTISALTALVT.....FLWPKYPRKNSFYLTAKA 111

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : A.Geneseq_032802.*

```
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT:*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT:*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT:*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT:*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT:*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT:*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT:*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT:*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT:*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT:*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT:*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT:*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT:*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT:*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT:*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT:*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT:*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT:*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT:*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT:*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT:*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	558	100.0	111	22	AAB60640
2	78	14.0	300	21	AAW81582
3	74	13.3	112	21	AAH21107
4	74	13.3	112	21	AAV93269
5	69.5	12.5	334	21	AAG31852
6	69.5	12.5	335	21	AAG31851
7	69.5	12.5	337	21	AAG31850
8	68	12.2	322	21	AAG11124
9	68	12.2	323	21	AAG11123
10	68	12.2	325	21	AAG11122
11	67.5	12.1	151	17	AAW95910

12	66.5	11.9	222	22	ABG02873	Novel human diageno
13	66	11.8	421	22	AAW39816	Human polypeptide
14	66	11.8	475	22	AAW41602	Human polypeptide
15	64.5	11.6	281	22	AAE04356	Oryza sativa rdtlf
16	64.5	11.6	564	22	AAE04359	Glycine max aspart
17	64.5	11.6	2785	21	AAV57148	Human down-regulat
18	64	11.5	970	22	AAW67548	Amino acid sequenc
19	63.5	11.4	295	22	AAW92147	C glutamicam prote
20	63.5	11.4	296	22	AAW9614	Corynebacterium gl
21	63.5	11.4	346	11	AAW06323	Human DNase. Homo
22	63.5	11.4	389	19	AAW96234	H. pylori GHP0 538
23	63.5	11.4	429	21	AAW54261	Arabidopsis thalia
24	63.5	11.4	430	21	AAW54260	Arabidopsis thalia
25	63.5	11.4	453	21	AAW54259	Arabidopsis thalia
26	63	11.3	196	22	AAW63940	Human prostate can
27	63	11.3	254	14	AAW33076	HpaI restriction e
28	63	11.3	981	20	AAW34476	Porphyromonas ging
29	63	11.3	985	20	AAW34352	Porphyromonas ging
30	63	11.3	1222	21	AAW36453	Arabidopsis thalia
31	63	11.3	1257	21	AAW36452	Arabidopsis thalia
32	63	11.3	1275	21	AAW36451	Arabidopsis thalia
33	62.5	11.2	255	22	AAW91022	C glutamicum prote
34	62.5	11.2	363	22	AAW01104	Novel human diageno
35	62.5	11.2	782	22	AAW46720	ACNV DNA polymeras
36	62.5	11.2	999	22	AAW01105	Novel human diageno
37	62	11.1	326	21	AAW29456	Arabidopsis thalia
38	62	11.1	330	20	AAW35229	Chlamydia pneumoni
39	62	11.1	330	21	AAW29455	Arabidopsis thalia
40	61.5	11.0	214	22	AAW71901	Arabidopsis thalia
41	61.5	11.0	465	22	AAW66746	Drosophila melanog
42	61.5	11.0	787	22	AAW58186	Drosophila melanog
43	61.5	11.0	787	22	AAW66743	Drosophila melanog
44	61.5	11.0	787	22	AAW66744	Drosophila melanog
45	61.5	11.0	787	22	AAW66745	Drosophila melanog

ALIGNMENTS

RESULT 1	AAW60640	standard; Protein: 111 AA.
ID	AAW60640	
AC	AAW60640	
XX	02-MAY-2001	(first entry)
DE	Moraxella catarrhalis strain ATCC43617 BASB122 protein.	
XX		
KW	BASB122 protein; strain ATCC43617; antigen; antibody; vaccine;	
KW	genetic immunisation; infection; upper respiratory tract; otitis media;	
KW	hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;	
KW	invasive disease; antibacterial; auditory.	
XX		
OS	Moraxella catarrhalis.	
XX		
PN	WO200109337-A2.	
PD	08-FEB-2001.	
XX		
PF	31-JUL-2000; 2000WO-EP07365.	
XX		
PR	30-JUL-1999; 99GB-0018034.	
PR	30-JUL-1999; 99GB-0018036.	
PA	(SMK) SMITHKLINE BEECHAM BIOLOGICALS.	
PI	Thomson J.	
XX		
DR	WPI: 2001-159874/16.	
XX	N-PSDB: AA59778.	
PT	New BASB122 and BASB124 polypeptides and polynucleotides from Moraxella	

Approved - Approved
Aug

PT catarrhalis strain ATCC 43617, useful as therapeutic agents or vaccines
 PT against bacterial infections, e.g. otitis media or pneumonia
 XX
 PS Claim 4; Page 66; 75pp; English.
 XX
 CC The invention relates to the Moraxella catarrhalis strain ATCC43617
 CC BASB122 and BASB124 proteins (AAB60640 and AAB60641, respectively)
 CC and to DNA encoding them (AAFS9778 and AAFS9779, respectively). The
 CC invention also relates to immunogenic fragments of the BASB122 and
 CC BASB124 proteins, expression vectors and host cells comprising BASB122
 CC or BASB124 nucleic acids, the recombinant production of BASB122 or
 CC BASB124, vaccine compositions comprising the BASB122 or BASB124 proteins
 CC or nucleic acids, an antibody against BASB122 or BASB124, therapeutic
 CC compositions comprising an anti-BASB122 or anti-BASB124 antibody, and a
 CC method of identifying a Moraxella catarrhalis infection via the
 CC detection of BASB122 or BASB124 proteins or antibodies. The vaccine
 CC compositions of the invention are useful as prophylactic or therapeutic
 CC agents against Moraxella catarrhalis infections in mammals, particularly
 CC humans. Moraxella catarrhalis is a Gram negative bacterium frequently
 CC isolated from the human upper respiratory tract, which is responsible for
 CC several pathological conditions. It is responsible for about 15% of
 CC otitis media cases in children (which can lead to temporary or permanent
 CC hearing loss). It also causes pneumonia in elderly people, and sinusitis,
 CC nosocomial infections and, less frequently, invasive diseases. BASB122 or
 CC BASB124 proteins or nucleotides may additionally be used in screening for
 CC novel antibacterial compounds, and in the diagnosis and staging of
 CC infections. The present sequence represents the Moraxella catarrhalis
 CC strain ATCC43617 BASB122 protein.
 XX
 SQ Sequence 111 AA;
 Query Match 100.0%; Score 558; DB 22; Length 111;
 Best Local Similarity 100.0%; Pred. No. 3.9e-61;
 Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MLNNHIRTISALTALIVGCVSTGNVAMKEONQRTIEQITIKGKTOKOISRSFGSAD 60
 DB 1 mlnhirltislaltalivgcvstgnvnmkeqngtqlkgtknkgsifsgsad 60
 QY 61 SISPMIVVIRKGFHTAIIAPNRMOELSLISFLWVKPYRPKNLSFYLAKA 111
 DB 61 sistmivvirkghfhtaiiapnrmoelstlisflwvkpyrpnklsfytaka 111
 RESULT 2
 AAY81582
 ID AAY81582 standard; Protein: 300 AA.
 AC AAY81582;
 XX
 DT 24-MAY-2000 (first entry)
 XX
 DE Streptococcus pneumoniae type 4 protein sequence #82.
 XX
 KW Streptococcus pneumoniae; vaccine; screening; protein antigen;
 KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
 KW pneumococcal disease.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN WO200006737-A2.
 PD 10-FEB-2000.
 XX
 PE 27-JUL-1999; 99WO-GB02451.
 XX
 PR 27-JUL-1998; 98GB-0016337.
 PR 19-MAR-1999; 99US-0125164.
 XX
 PA (MICR-) MICROBIAL TECHNICS LTD.
 XX
 PI Gilbert CRG, Hansbro PM;

XX
 DR WPI; 2000-195300/17.
 XX
 PT New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing
 PT or inhibiting expression of the protein
 PS Claim 1; Page 80; 108pp; English.
 XX
 CC AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from Streptococcus pneumoniae. AAO5407 to AAO5590 represent
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
 CC The sequences have antibacterial and antiinflammatory properties.
 CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonising, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. AAO5591 to AAO5614 represent primers used in the
 CC exemplification of the present invention.
 XX
 SQ Sequence 300 AA;
 Query Match 14.0%; Score 78; DB 21; Length 300;
 Best Local Similarity 28.3%; Pred. No. 0.36;
 Matches 36; Conservative 18; Mismatches 39; Indels 34; Gaps 6;
 QY 10 ISALLTA-LIVTGCVST--GNVAMKEONQRTIEQITII-----KGTN 48
 DB 78 lgalvtavilvtgsvllevtklllpqpvndegllwlgilaatlmlaaslvvgkgtk 137
 QY 49 KOEISSRFGSADSISEF-----MIWIKRGHHTAIIAPNRMOELSLISFLWVKPYRPKNL 103
 DB 138 neslshlfiedltgwwavilmavilrftdwylldp-----lslvlsffllskalpr-- 190
 QY 104 SFYLTAK 110
 DB 191 -fwstlk 196
 RESULT 3
 AAB21107
 ID AAB21107 standard; Protein: 112 AA.
 AC AAB21107;
 XX
 DT 19-JAN-2001 (first entry)
 XX
 DE Neisseria meningitidis BASB059 protein.
 XX
 KW BASB059: respiratory tract infection; invasive bacterial disease;
 KW bacteraemia; meningitis.
 XX
 OS Neisseria meningitidis.
 XX
 PN WO2000044904-A1.
 PD 03-AUG-2000.
 XX
 PE 25-JAN-2000; 2000WO-EP00561.
 XX
 PR 29-JAN-1999; 99GB-0002070.
 XX
 PA (SMIK) SMITHKLINE BEECHAM BIOLOGICALS.
 XX
 PI Thonnard J;
 XX
 DR WPI; 2000-505978/45.
 DR N-PSDB; AAA76029.

PT New isolated polypeptide from *Neisseria meningitidis* is useful for
detection and treatment of *N. meningitidis* infection -

PS Claim 3; Page 61; 77pp; English.

XX The present sequence is the *Neisseria meningitidis* BASB059 protein. The
CC protein, its gene, agonists and antagonists can be used in diagnostic
CC assays for the bacterium, as vaccines to prevent infection and as
CC treatments for bacterial infection, particularly those caused by
CC *Neisseria meningitidis*, such as upper respiratory tract infections,
CC invasive bacterial diseases, bacteraemia and meningitis.

XX Sequence 112 AA;

Query Match 13.3%; Score 74; DB 21; Length 112;
Best Local Similarity 31.2%; Pred. No. 0.3;
Matches 20; Conservative 10; Mismatches 34; Indels 0; Gaps 0;

OY 7 RLTIALTLALVTGCVSTGNVAMKEONQOTIEQTIKGTNKQKISRFSGADSISPMI 66
DB 4 RLVSAAFVALALAGCGSINNVTSDQKIGERAFAIYVGNVAKISNINFTFA 63

OY 67 VVIR 70
DB 64 tvrk 67

RESULT 4

AA93269
ID AAY93269 standard; Protein: 112 AA.

XX AAY93269;

DT 04-SEP-2000 (first entry)

XX Amino acid sequence of a polypeptide of a *Neisseria* pathogenic strain.

XX Pathogenic strain; *Neisseria*; vaccine; *Neisseria* infection.

XX *Neisseria meningitidis*.

XX WO200026375-A2.

XX 11-MAY-2000.

XX 28-OCT-1999; 99WO-FR02643.

XX 30-OCT-1998; 98FR-0013693.

PA (INNR) PASTEUR MERIEUX SERUMS & VACCINS SA.
(INNR) INSERM INST NAT SANTE & RECH MEDICALE.

PI Anjame L, Bouchardon A, Renaud-Mongenie G, Rokhl B, Nassif X;

PI Tinsley C., Perrin A;

DR WPI: 2000-365622/31.

DR N-PSDB; AAA15300.

XX New polypeptide vaccine for pathogenic *Neisseria* useful in therapeutic
PT or preventative vaccines and for diagnosis -

PS Claim 5; Page 69-70; 187pp; French.

XX The present sequence represents a protein that is specific for pathogenic
CC strains of *Neisseria*. The polynucleotides, polypeptides, or their
CC antigenic fragments, are used in vaccines to treat or protect against
CC *Neisseria* infections, particularly by *N. meningitidis*. The
CC polynucleotide sequence is also used for recombinant production of
CC the polypeptide and to produce attenuated *Neisseria* strains that
CC overexpress it, or express it in a non-toxic mutant form.

XX Sequence 112 AA;

Query Match 13.3%; Score 74; DB 21; Length 112;
Best Local Similarity 31.2%; Pred. No. 0.3;
Matches 20; Conservative 10; Mismatches 34; Indels 0; Gaps 0;

OY 7 RLTIALTLALVTGCVSTGNVAMKEONQOTIEQTIKGTNKQKISRFSGADSISPMI 66
DB 4 RLVSAAFVALALAGCGSINNVTSDQKIGERAFAIYVGNVAKISNINFTFA 63

OY 67 VVIR 70
DB 64 tvrk 67

RESULT 5

AA631852
ID AAG31852 standard; Protein: 334 AA.

XX AAG31852;

DT 17-OCT-2000 (first entry)

XX Arabidopsis thaliana protein fragment SEQ ID NO: 38321.

XX Protein identification; signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.

XX Arabidopsis thaliana.

XX EPI033405-A2.

XX 06-SEP-2000.

XX 25-FEB-2000; 2000EP-0301439.

XX 25-FEB-1999; 99US-0121825.

XX 05-MAR-1999; 99US-0123180.

XX 09-MAR-1999; 99US-0123548.

XX 23-MAR-1999; 99US-0125788.

XX 25-MAR-1999; 99US-0126264.

XX 29-MAR-1999; 99US-0126785.

XX 01-APR-1999; 99US-0127462.

XX 06-APR-1999; 99US-0128234.

XX 08-APR-1999; 99US-0128714.

XX 16-APR-1999; 99US-0129845.

XX 19-APR-1999; 99US-0130077.

XX 21-APR-1999; 99US-0130449.

XX 23-APR-1999; 99US-0130510.

XX 28-APR-1999; 99US-0130891.

XX 30-APR-1999; 99US-0131449.

XX 04-MAY-1999; 99US-0132407.

XX 05-MAY-1999; 99US-0132484.

XX 06-MAY-1999; 99US-0132485.

XX 06-MAY-1999; 99US-0132486.

XX 07-MAY-1999; 99US-0132487.

XX 11-MAY-1999; 99US-0132486.

XX 14-MAY-1999; 99US-0134218.

XX 14-MAY-1999; 99US-0134219.

XX 14-MAY-1999; 99US-0134221.

XX 14-MAY-1999; 99US-0134370.

XX 18-MAY-1999; 99US-0134768.

XX 19-MAY-1999; 99US-0134941.

XX 20-MAY-1999; 99US-0135124.

XX 21-MAY-1999; 99US-0135353.

XX 24-MAY-1999; 99US-0135629.

XX 25-MAY-1999; 99US-0136021.

XX 27-MAY-1999; 99US-0136392.

XX 28-MAY-1999; 99US-0136782.

XX 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139753.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139839.
PR 23-JUN-1999; 99US-0140353.
PR 24-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140981.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144684.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149358.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160981.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match

12.5%; Score 69.5; DB 21; Length 334;

PR	05-AUG-1999;	99US-0147192.
PR	05-AUG-1999;	99US-0147260.
PR	06-AUG-1999;	99US-0147303.
PR	06-AUG-1999;	99US-0147416.
PR	09-AUG-1999;	99US-0147493.
PR	09-AUG-1999;	99US-0147935.
PR	10-AUG-1999;	99US-0148171.
PR	11-AUG-1999;	99US-0148319.
PR	12-AUG-1999;	99US-0148341.
PR	13-AUG-1999;	99US-0148565.
PR	16-AUG-1999;	99US-0148684.
PR	17-AUG-1999;	99US-0149175.
PR	18-AUG-1999;	99US-0149426.
PR	20-AUG-1999;	99US-0149722.
PR	20-AUG-1999;	99US-0149723.
PR	20-AUG-1999;	99US-0149929.
PR	23-AUG-1999;	99US-0149930.
PR	23-AUG-1999;	99US-0150066.
PR	25-AUG-1999;	99US-0150864.
PR	26-AUG-1999;	99US-0150884.
PR	27-AUG-1999;	99US-0151065.
PR	27-AUG-1999;	99US-0151066.
PR	27-AUG-1999;	99US-0151080.
PR	30-AUG-1999;	99US-0151303.
PR	31-AUG-1999;	99US-0151438.
PR	01-SEP-1999;	99US-0151930.
PR	07-SEP-1999;	99US-0152363.
PR	10-SEP-1999;	99US-0153070.
PR	13-SEP-1999;	99US-0153758.
PR	15-SEP-1999;	99US-0154018.
PR	16-SEP-1999;	99US-0154039.
PR	20-SEP-1999;	99US-0154779.
PR	22-SEP-1999;	99US-0155139.
PR	23-SEP-1999;	99US-0155486.
PR	24-SEP-1999;	99US-0155659.
PR	28-SEP-1999;	99US-0156458.
PR	29-SEP-1999;	99US-0156596.
PR	04-OCT-1999;	99US-0157117.
PR	05-OCT-1999;	99US-0157753.
PR	06-OCT-1999;	99US-0157865.
PR	07-OCT-1999;	99US-0158029.
PR	08-OCT-1999;	99US-0158232.
PR	12-OCT-1999;	99US-0158369.
PR	13-OCT-1999;	99US-0159283.
PR	13-OCT-1999;	99US-0159284.
PR	13-OCT-1999;	99US-0159295.
PR	14-OCT-1999;	99US-0159329.
PR	14-OCT-1999;	99US-0159330.
PR	14-OCT-1999;	99US-0159331.
PR	14-OCT-1999;	99US-0159637.
PR	14-OCT-1999;	99US-0159638.
PR	18-OCT-1999;	99US-0159584.
PR	18-OCT-1999;	99US-0159584.
PR	21-OCT-1999;	99US-0160741.
PR	21-OCT-1999;	99US-0160767.
PR	21-OCT-1999;	99US-0160770.
PR	21-OCT-1999;	99US-0160776.
PR	21-OCT-1999;	99US-0160814.
PR	21-OCT-1999;	99US-0160815.
PR	22-OCT-1999;	99US-0160980.
PR	22-OCT-1999;	99US-0160981.
PR	22-OCT-1999;	99US-0160989.
PR	25-OCT-1999;	99US-0161404.
PR	25-OCT-1999;	99US-0161405.
PR	25-OCT-1999;	99US-0161406.
PR	26-OCT-1999;	99US-0161359.
PR	26-OCT-1999;	99US-0161360.
PR	26-OCT-1999;	99US-0161361.
PR	28-OCT-1999;	99US-0161920.
PR	28-OCT-1999;	99US-0161982.
PR	28-OCT-1999;	99US-0161993.
PR	29-OCT-1999;	99US-0162142.

PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137722.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 21-JUN-1999;	99US-0139763.
PR 22-JUN-1999;	99US-0139817.
PR 23-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140693.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142053.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 23-JUL-1999;	99US-0145224.
PR 26-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147935.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 16-AUG-1999;	99US-0148684.
PR 17-AUG-1999;	99US-0149175.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0152070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154179.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156596.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0159293.
PR 13-OCT-1999;	99US-0159294.
PR 13-OCT-1999;	99US-0159295.
PR 14-OCT-1999;	99US-0159329.
PR 14-OCT-1999;	99US-0159330.
PR 14-OCT-1999;	99US-0159331.
PR 14-OCT-1999;	99US-0159637.
PR 14-OCT-1999;	99US-0159638.
PR 18-OCT-1999;	99US-0159584.
PR 21-OCT-1999;	99US-0160741.
PR 21-OCT-1999;	99US-0160767.
PR 21-OCT-1999;	99US-0160768.
PR 21-OCT-1999;	99US-0160770.
PR 21-OCT-1999;	99US-0160814.
PR 21-OCT-1999;	99US-0160815.
PR 22-OCT-1999;	99US-0160980.
PR 22-OCT-1999;	99US-0160981.
PR 22-OCT-1999;	99US-0160989.
PR 23-OCT-1999;	99US-0161404.
PR 25-OCT-1999;	99US-0161405.
PR 25-OCT-1999;	99US-0161406.
PR 26-OCT-1999;	99US-0161359.
PR 26-OCT-1999;	99US-0161360.
PR 26-OCT-1999;	99US-0161361.
PR 28-OCT-1999;	99US-0161920.
PR 28-OCT-1999;	99US-0161992.
PR 28-OCT-1999;	99US-0161993.
PR 29-OCT-1999;	99US-0162142.

PR	27-MAY-1999;	9905-0136392.
PR	28-MAY-1999;	9905-0136782.
PR	01-JUN-1999;	9905-0137222.
PR	03-JUN-1999;	9905-0137528.
PR	04-JUN-1999;	9905-0137502.
PR	07-JUN-1999;	9905-0137724.
PR	08-JUN-1999;	9905-0138094.
PR	10-JUN-1999;	9905-0138540.
PR	14-JUN-1999;	9905-0139119.
PR	16-JUN-1999;	9905-0139432.
PR	16-JUN-1999;	9905-0139452.
PR	17-JUN-1999;	9905-0139424.
PR	18-JUN-1999;	9905-0139455.
PR	18-JUN-1999;	9905-0139455.
PR	18-JUN-1999;	9905-0139457.
PR	18-JUN-1999;	9905-0139458.
PR	18-JUN-1999;	9905-0138459.
PR	18-JUN-1999;	9905-0138460.
PR	18-JUN-1999;	9905-0138461.
PR	18-JUN-1999;	9905-0139462.
PR	18-JUN-1999;	9905-0139463.
PR	18-JUN-1999;	9905-0139750.
PR	18-JUN-1999;	9905-0139763.
PR	21-JUN-1999;	9905-0139817.
PR	22-JUN-1999;	9905-0139899.
PR	23-JUN-1999;	9905-0140353.
PR	23-JUN-1999;	9905-0140354.
PR	24-JUN-1999;	9905-0140695.
PR	28-JUN-1999;	9905-0140823.
PR	28-JUN-1999;	9905-0140991.
PR	30-JUN-1999;	9905-0141287.
PR	01-JUL-1999;	9905-0141842.
PR	02-JUL-1999;	9905-0142055.
PR	06-JUL-1999;	9905-0142390.
PR	08-JUL-1999;	9905-0142803.
PR	09-JUL-1999;	9905-0142920.
PR	12-JUL-1999;	9905-0142977.
PR	13-JUL-1999;	9905-0143542.
PR	14-JUL-1999;	9905-0143654.
PR	15-JUL-1999;	9905-0144005.
PR	16-JUL-1999;	9905-0144085.
PR	16-JUL-1999;	9905-0144086.
PR	19-JUL-1999;	9905-0144331.
PR	19-JUL-1999;	9905-0144332.
PR	19-JUL-1999;	9905-0144334.
PR	20-JUL-1999;	9905-0144335.
PR	20-JUL-1999;	9905-0144352.
PR	21-JUL-1999;	9905-0144884.
PR	21-JUL-1999;	9905-0144814.
PR	21-JUL-1999;	9905-0145086.
PR	21-JUL-1999;	9905-0145087.
PR	22-JUL-1999;	9905-0145087.
PR	22-JUL-1999;	9905-0145087.
PR	22-JUL-1999;	9905-0145199.
PR	22-JUL-1999;	9905-0145145.
PR	23-JUL-1999;	9905-0145216.
PR	23-JUL-1999;	9905-0145224.
PR	26-JUL-1999;	9905-0145276.
PR	27-JUL-1999;	9905-0145913.
PR	27-JUL-1999;	9905-0145918.
PR	27-JUL-1999;	9905-0145918.
PR	28-JUL-1999;	9905-0145951.
PR	02-AUG-1999;	9905-0146388.
PR	02-AUG-1999;	9905-0146587.
PR	03-AUG-1999;	9905-0147038.

PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0143542.
PR 13-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 19-JUL-1999; 99US-0144336.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 26-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149375.
PR 18-AUG-1999; 99US-0149426.
PR 18-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150884.
PR 26-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 29-SEP-1999; 99US-0157117.
PR 04-OCT-1999; 99US-0157753.
PR 05-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158332.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 12.2%; Score 68; DB 21; Length 323;
Best Local Similarity 28.8%; Pred. No. 6.9';
Matches 17; Conservative 18; Mismatches 16; Indels 8; Gaps 3;
OY 3 NHHIRITISALT---ALLVTGCVS-TGNVA--KKEONQRIEFTIIKGTNKOEIS 53

Db 59 nrtsvtsaateplvltvgagrtgqlvykkkterseqfvarglvrtkeskekin 117

RESULT 10

AG11122

ID AG11122 standard; Protein; 325 AA.

AC AG11122;

XX 17-OCT-2000 (first entry)

DT

XX Arabidopsis thaliana protein fragment SEQ ID NO: 9721.

XX

KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

XX Arabidopsis thaliana.

OS

XX EPI033405-A2.

PN

XX 06-SEP-2000.

PD

XX

XX 25-FEB-2000; 2000EP-0301439.

PE

XX

XX 25-FEB-1999; 99US-0121825.

PR 05-MAR-1999; 99US-0123180.

PR 09-MAR-1999; 99US-0123348.

PR 23-MAR-1999; 99US-0125788.

PR 25-MAR-1999; 99US-0126264.

PR 29-MAR-1999; 99US-0126785.

PR 01-APR-1999; 99US-0127462.

PR 06-APR-1999; 99US-0128234.

PR 08-APR-1999; 99US-0128714.

PR 16-APR-1999; 99US-0129845.

PR 19-APR-1999; 99US-0130077.

PR 21-APR-1999; 99US-0130449.

PR 23-APR-1999; 99US-0130510.

PR 23-APR-1999; 99US-0130891.

PR 28-APR-1999; 99US-0131449.

PR 30-APR-1999; 99US-0132048.

PR 30-APR-1999; 99US-0132407.

PR 04-MAY-1999; 99US-0132484.

PR 05-MAY-1999; 99US-0132485.

PR 06-MAY-1999; 99US-0132486.

PR 06-MAY-1999; 99US-0132487.

PR 07-MAY-1999; 99US-0132863.

PR 11-MAY-1999; 99US-0134256.

PR 14-MAY-1999; 99US-0134218.

PR 14-MAY-1999; 99US-0134219.

PR 14-MAY-1999; 99US-0134221.

PR 14-MAY-1999; 99US-0134370.

PR 18-MAY-1999; 99US-0134768.

PR 19-MAY-1999; 99US-0134941.

PR 20-MAY-1999; 99US-0135124.

PR 21-MAY-1999; 99US-0135353.

PR 24-MAY-1999; 99US-0135629.

PR 25-MAY-1999; 99US-0136021.

PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.

PR 01-JUN-1999; 99US-0137222.

PR 03-JUN-1999; 99US-0137528.

PR 04-JUN-1999; 99US-0137502.

PR 07-JUN-1999; 99US-0137724.

PR 08-JUN-1999; 99US-0138094.

PR 10-JUN-1999; 99US-0138540.

PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.

PR 16-JUN-1999; 99US-0139452.

PR 16-JUN-1999; 99US-0139453.

PR 17-JUN-1999; 99US-0139492.

PR 18-JUN-1999; 99US-0139454.

PR 18-JUN-1999; 99US-0139455.

PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.

PR 18-JUN-1999; 99US-0139458.

PR 18-JUN-1999; 99US-0139459.

PR 18-JUN-1999; 99US-0139460.

PR 18-JUN-1999; 99US-0139461.

PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.

PR 18-JUN-1999; 99US-0139750.

PR 18-JUN-1999; 99US-0139763.

PR 21-JUN-1999; 99US-0139817.

PR 22-JUN-1999; 99US-0139891.

PR 23-JUN-1999; 99US-0140353.

PR 23-JUN-1999; 99US-0140354.

PR 24-JUN-1999; 99US-0140355.

PR 24-JUN-1999; 99US-0140695.

PR 28-JUN-1999; 99US-0139750.

PR 29-JUN-1999; 99US-0140823.

PR 30-JUN-1999; 99US-0140991.

PR 01-JUL-1999; 99US-0141287.

PR 01-JUL-1999; 99US-0141842.

PR 01-JUL-1999; 99US-0142154.

PR 02-JUL-1999; 99US-0142055.

PR 06-JUL-1999; 99US-0142390.

PR 08-JUL-1999; 99US-0142803.

PR 09-JUL-1999; 99US-0142920.

PR 12-JUL-1999; 99US-0142977.

PR 13-JUL-1999; 99US-0143542.

PR 14-JUL-1999; 99US-0143624.

PR 15-JUL-1999; 99US-0144005.

PR 16-JUL-1999; 99US-0144085.

PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.

PR 19-JUL-1999; 99US-0144331.

PR 19-JUL-1999; 99US-0144332.

PR 19-JUL-1999; 99US-0144333.

PR 19-JUL-1999; 99US-0144334.

PR 19-JUL-1999; 99US-0144335.

PR 20-JUL-1999; 99US-0144352.

PR 20-JUL-1999; 99US-0144632.

PR 20-JUL-1999; 99US-0144884.

PR 21-JUL-1999; 99US-0144814.

PR 21-JUL-1999; 99US-0145086.

PR 22-JUL-1999; 99US-0145088.

PR 22-JUL-1999; 99US-0145085.

PR 22-JUL-1999; 99US-0145087.

PR 22-JUL-1999; 99US-0145089.

PR 22-JUL-1999; 99US-0145192.

PR 23-JUL-1999; 99US-0145145.

PR 23-JUL-1999; 99US-0145218.

PR 23-JUL-1999; 99US-0145224.

PR 26-JUL-1999; 99US-0145276.

PR 27-JUL-1999; 99US-0145913.

PR 27-JUL-1999; 99US-0145918.

PR 27-JUL-1999; 99US-0145919.

PR 28-JUL-1999; 99US-0145951.

PR 02-AUG-1999; 99US-0146386.

PR 02-AUG-1999; 99US-0146388.

PR 02-AUG-1999; 99US-0146389.

PR 03-AUG-1999; 99US-0147038.

PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.

PR 05-AUG-1999; 99US-0147192.

PR 05-AUG-1999; 99US-0147260.

PR 06-AUG-1999; 99US-0147302.

PR 06-AUG-1999; 99US-0147416.

PR 09-AUG-1999; 99US-0147493.

PR 09-AUG-1999; 99US-0147935.

PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.

PR 12-AUG-1999; 99US-0148341.

PR 13-AUG-1999; 99US-0148565.

PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.

```

PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158393.
PR 13-OCT-1999; 99US-0158993.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160776.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 25-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

Query Match 12.2%; Score 68; DB 21; Length 325;
 Best Local Similarity 28.8%; Pred. NO. 6.9;
 Matches 17; Conservative 18; Mismatches 16; Indels 8; Gaps 3;

```

OY 3 NHIRLTLSALLT---ALIVTGCVS-TGNVA---MKRONOQTIOTIKTKTKNOEIS 53
    | :|:| | | :|:| | | | :|:|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 61 nrtvsvtsaatleptlvltgaggrtgivkykliersegfvarglvrtkeskekin 119

```

RESULT 11
 AAR95910.
 ID AAR95910 standard; Protein; 151 AA.

```

XX AC AAR95910;
XX DX 17-AUG-1996 (first entry)
XX DX Opacity associated protein Oapb.
XX DE Opacity associated protein; OapA; OapB; vector; vaccine; antibody;
XX KW diagnosis; immunogen.
XX OS Haemophilus influenzae strain Rd variant H175.
XX PN W09613515-A1.
XX PD 09-MAY-1996.
XX PF 24-OCT-1995; 95WO-US13672.
XX FR 31-OCT-1994; 94US-0332576.
XX PA (CHIL-) CHILDRENS HOSPITAL PHILADELPHIA.
XX PI Weiser JN;
XX DR WPI; 1996-239445/24.
XX DR N-PSDB; AAT27729.
XX PT New opacity associated protein of Haemophilus influenzae - and
XX PT related DNA, vectors, transformed hosts and antibodies, useful in
XX PT vaccines and for diagnosis.
XX PS Claim 3; Fig 1a-c; 48pp; English.
XX PS Hemophilus influenzae opacity associated proteins OapA (AAR95909)
XX CC and OapB (AAR95910) are membrane proteins useful as immunogens in
XX CC and OapB (AAR95910) are membrane proteins useful as immunogens in
XX CC vaccines against typhoid and non-typhoid H. influenzae infection.
XX CC They were identified as the products of 2 open reading frames of
XX CC an oap gene (AAT27729) isolated from the spontaneous opacity variant
XX CC H175. Recombinant Oap proteins can be expressed in host cells
XX CC carrying a vector including an oap gene sequence. Antibodies
XX CC raised against the proteins are useful for detecting H. influenzae
XX CC or for passive immunisation.
XX CC
XX SQ Sequence 151 AA;

```

Query Match 12.1%; Score 67.5; DB 17; Length 151;
 Best Local Similarity 33.3%; Pred. NO. 2.9; Indels 3; Gaps 1;
 Matches 14; Conservative 10; Mismatches 15

```

OY 11 SALLTALVTGCVSTGNVAMKRONOQTIET--QTIIKGTNK 49
    | :|:| | | | | :|:| :|:| :|:|:| :|:|:| :|:|:| :|:|:|
Db 6 sltfallimtcgvnqanvtltpgaqkmqevkdalqgseadr 47

```

RESULT 12
 ABG02873
 ID ABG02873 standard; Protein; 222 AA.
 AC ABG02873;
 DX 13-FEB-2002 (first entry)
 DE Novel human diagnostic protein #2864.
 DX
 KW Human; chromosome mapping; gene mapping; gene therapy; forensic;
 KW food supplement; medical imaging; diagnostic; genetic disorder.
 OS Homo sapiens.
 XX WO200175067-A2.
 PN
 PD 11-OCT-2001.

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:35:51 ; Search time 12.91 Seconds
(without alignments)
210.011 Million cell updates/sec

Title: US-10-048-197-2

Perfect score: 558

Sequence: 1 MLNHHIRLITSLALTLALVT.....FLWKYRPRKNSFLYLAKA 111

Scoring table: BLOSUM62

Searched: Gapop 10.0 , Gapext 0.5

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCtus_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	Query	Score	Match	Length	DB	ID	Description
No.							
1	67.5	12.1	151	1	US-08-332-576-3		Sequence 3, Appl1
2	67.5	12.1	151	5	PCT-US95-13672-3		Sequence 3, Appl1
3	63.5	11.4	346	1	US-07-895-300A-1		Sequence 1, Appl1
4	63.5	11.4	346	1	US-08-458-367-1		Sequence 1, Appl1
5	63.5	11.4	346	5	PCT-US93-05136-1		Sequence 1, Appl1
6	59.5	10.7	399	3	US-08-651-999A-3		Sequence 3, Appl1
7	59.5	10.7	399	4	US-09-385-752-3		Sequence 3, Appl1
8	58.5	10.5	229	4	US-09-247-373B-48		Sequence 48, Appl1
9	58.5	10.5	246	3	US-09-238-481-2		Sequence 2, Appl1
10	57	10.2	150	1	US-08-374-843B-6		Sequence 6, Appl1
11	57	10.2	150	1	US-08-374-843B-10		Sequence 10, Appl1
12	57	10.2	150	2	US-08-905-420-6		Sequence 6, Appl1
13	57	10.2	150	2	US-08-905-420-10		Sequence 10, Appl1
14	57	10.2	193	1	US-08-017-114-19		Sequence 19, Appl1
15	57	10.2	193	3	US-08-505-307-19		Sequence 19, Appl1
16	57	10.2	193	4	US-09-082-920-4		Sequence 4, Appl1
17	57	10.2	193	5	PCT-US94-02034-19		Sequence 19, Appl1
18	57	10.2	1817	4	US-09-004-838-125		Sequence 125, App
19	57	10.2	3079	5	PCT-US94-00198-4		Sequence 8, Appl1
20	56.5	10.1	193	1	US-08-564-972-8		Sequence 27, Appl1
21	56.5	10.1	194	1	US-08-148-058A-29		Sequence 27, Appl1
22	56.5	10.1	194	1	US-08-148-058A-27		Sequence 27, Appl1
23	56.5	10.1	194	1	US-08-478-042-29		Sequence 29, Appl1
24	56.5	10.1	194	1	US-08-478-042-29		Sequence 29, Appl1
25	56.5	10.1	194	2	US-08-645-215-27		Sequence 27, Appl1
26	56.5	10.1	194	2	US-08-645-215-29		Sequence 29, Appl1
27	56.5	10.1	194	2	US-08-466-604-27		Sequence 27, Appl1

28	56.5	10.1	194	2	US-08-466-604-29	Sequence 29, Appl1
29	56.5	10.1	264	1	US-08-107-042-2	Sequence 2, Appl1
30	56.5	10.1	402	3	US-08-948-997-4	Sequence 4, Appl1
31	56.5	10.1	402	4	US-09-348-817A-4	Sequence 4, Appl1
32	56.5	10.1	535	2	US-08-564-972-1	Sequence 1, Appl1
33	56.5	10.1	535	4	US-09-171-969-9	Sequence 9, Appl1
34	56.5	10.1	3289	2	US-08-477-451-2	Sequence 2, Appl1
35	56	10.0	129	4	US-09-476-482-6	Sequence 6, Appl1
36	56	10.0	2020	1	US-07-551-531-2	Sequence 2, Appl1
37	56	10.0	2485	5	PCT-US94-00198-1	Sequence 1, Appl1
38	56	10.0	2485	5	PCT-US94-00198-2	Sequence 1, Appl1
39	56	10.0	2818	1	US-08-510-284-1	Sequence 1, Appl1
40	56	10.0	2818	1	US-08-411-389-2	Sequence 2, Appl1
41	56	10.0	2818	4	US-09-542-331-2	Sequence 2, Appl1
42	55.5	9.9	217	4	US-09-248-588-9	Sequence 9, Appl1
43	55.5	9.9	379	3	US-08-840-204-3	Sequence 3, Appl1
44	55.5	9.9	390	1	US-08-121-714-6	Sequence 6, Appl1
45	55.5	9.9	390	1	US-08-477-108A-6	Sequence 6, Appl1

ALIGNMENTS

```
RESULT 1
US-08-332-576-3
; Sequence 3, Application US/08332576
; Patent No. 5756105
; GENERAL INFORMATION:
; APPLICANT: Weiser, Jeffrey M.
; TITLE OF INVENTION: Vaccines for Haemophilus Influenza
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESS: Woodcock Washburn Kurtz Mackiewicz and No. 5756105rls
; STREET: One Liberty Place - 46th Floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: U.S.A.
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/332,576
; FILING DATE:
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Ralph, Rebecca L.
; REGISTRATION NUMBER: 35,152
; REFERENCE/DOCKET NUMBER: CH-536
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215-568-3439
; TELEFAX: 215-568-3439
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 151 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-332-576-3

Query Match 12.1%; Score 67.5; DB 1; Length 151;
Best Local Similarity 33.3%; Pred. No. 0.31;
Matches 14; Conservative 10; Mismatches 15; Indels 3; Gaps 1;

QY 11 SALTALVTGCGVSGNVAKEEONQOTIE--QTIRKGTNK 49
I : ||||:|||| || : : : ||: : :
DB 6 SLFTALMTGCVQANANTTTPQAQKQVEKVDALQKGEADR 47
RESUL 2
```

[illegible]

4
 US-08-458-367-1
 ; Sequence 1, Application US/08458367
 ; Patent No. 5783433
 ;
 GENERAL INFORMATION:
 ;
 APPLICANT: Frenzy, John
 ;
 APPLICANT: Shire, Steven J.
 ;
 APPLICANT: Sliwowski, Mary B.
 ;
 TITLE OF INVENTION: PURIFIED FORMS OF DNase
 ;
 NUMBER OF SEQUENCES: 18
 ;
 CORRESPONDENCE ADDRESS:
 ;
 ADDRESSEE: Genentech, Inc.
 ;
 STREET: 460 Point San Bruno Blvd
 ;
 CITY: South San Francisco
 ;
 STATE: California
 ;
 COUNTRY: USA
 ;
 ZIP: 94080
 ;
 COMPUTER READABLE FORM:
 ;
 MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
 ;
 COMPUTER: IBM PC compatible
 ;
 OPERATING SYSTEM: PC-DOS/MS-DOS
 ;
 SOFTWARE: WinPaln (Genentech)
 ;
 CURRENT APPLICATION DATA:
 ;
 APPLICATION NUMBER: US/08/458,367
 ;
 FILING DATE: 02-Jun-1995
 ;
 CLASSIFICATION: 435
 ;
 PRIOR APPLICATION DATA:
 ;
 APPLICATION NUMBER: 08/409631
 ;
 FILING DATE: 22-Mar-1995
 ;
 PRIOR APPLICATION DATA:
 ;
 APPLICATION NUMBER: 08/348284
 ;
 FILING DATE: 30-No. 5783433-1994
 ;
 PRIOR APPLICATION DATA:

```

;      1EEN:  210/2/1  7400
;      INFORMATION FOR SEQ ID NO: 1:

```

US-08-651-999A-3

Query Match	10.78;	Score	59.5;	DB	3;	Length	359;
Best Local Similarity	23.98;	Pred. No.	14;				
Matches	26;	Conservative	14;	Mismatches	40;	Indels	29;
						Gaps	3;

```

QY      12  ALLTALLVTGCVSTGNAMKEQNOOTIEQ-----TIKGTKTKEISSFPGSAS----- 61
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db      79  SLATALLASVDGNPAIYGRRAIAMDKVSRSRGIGTIOVQTKSEITITFLNLSERWID 138
QY      62  -----ISMIVY---IKGHTALLANPQOELSLIIS 91
          :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
Db     139  DHTRAVITIEFSAMQAINIEFSVQVLLVETIPKSGIYLPNSWESVRLIKS 187

```

RESULT 7
 US-09-385-752-3
 Sequence 3, Application us/09385752
 Patent No. 6228591
 GENERAL INFORMATION:
 APPLICANT: Stefan Somlo and Toshio Mochizuki
 TITLE OF INVENTION: POLYCYSTIC KIDNEY DISEASE PKD2 GENE
 TITLE OF INVENTION: AND USES THEREOF
 NUMBER OF SEQUENCES: 12
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: AMSTER, ROTHSTEIN & EBENSTEIN
 STREET: 90 PARK AVENUE
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: U.S.A.
 ZIP: 10016
 COMPUTER READABLE FORM:
 MEDIUM TYPE: 3.5 INCH 1.44 Mb STORAGE
 MEDIUM TYPE: DISKETTE
 COMPUTER: IBM PC COMPATIBLE
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: ASCII
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/385,752
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/651,999
 FILING DATE: MAY 23, 1996
 ATTORNEY/AGENT INFORMATION:
 NAME: ELIZABETH A. BOGOSTAN
 REGISTRATION NUMBER: 39,911
 REFERENCE/DOCKET NUMBER: 96700/395
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 697-5995
 TELEFAX: (212) 286-0854 or 286-0082
 TELEX: TWX 710-581-4766
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 399 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 DESCRIPTION: NO
 HYPOTHETICAL: NO
 FRAGMENT TYPE: internal fragment

```

US-09-247-373B-48
:
: Sequence 48, Application US/09247373B
: Patent No. 6168954
:
: GENERAL INFORMATION:
:
: APPLICANT: MCGONIGLE, BRIAN
:
: APPLICANT: O'KEEFE, DANIEL
:
: TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
:
: FILE REFERENCE: CL-1108-A
:
: CURRENT APPLICATION NUMBER: US/09/247,373B
:
: CURRENT FILING DATE: 1999-02-10
:
: PRIOR APPLICATION NUMBER: 08/924,747
:
: PRIOR FILING DATE: 1997-09-05
:
: NUMBER OF SEQ ID NOS: 56
:
: SOFTWARE: Microsoft Office 97
:
: SEQ ID NO: 48
:
: LENGTH: 229
:
: TYPE: PRT
:
: ORGANISM: SOYBEAN
:
US-09-247-373B-48

```

```

RESULT          9
US-09-238-481-2
; Sequence 2, Application US/09238481
; Patent No. 6110704
; GENERAL INFORMATION:
; APPLICANT: Huang, Jianzhong
; APPLICANT: McDevitt, Damien
; TITLE OF INVENTION: Fabg
; FILE REFERENCE: G610192
; CURRENT APPLICATION NUMBER: US/09/238,481
; CURRENT FILING DATE: 1999-01-28
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 246
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-238-481-2

```

RESULT 10
US-08-374-843B-6
; Sequence 6, Application US/08374843B
; Patent No. 5726016
GENERAL INFORMATION:
; APPLICANT: Demuth, Donald R.
;


```

; Sequence 19, Application 05/0000000
; Patent No. 6083691
; GENERAL INFORMATION:
;

```

Query Match	10.2%;	Score 57;	DB 2;	Length 150
Best Local Similarity	27.0%;	Pred. No. 7.6;		

```

; APPLICANT: Deretic, Vojo
; APPLICANT: Martin, Daniel W.
; TITLE OF INVENTION: DETECTION OF CONVERSION TO MUCCOYD IN
; TITLE OF INVENTION: PSEUDOMONAS AERUGINOSA INFECTING CYSTIC FIBROSIS PATIENTS
; NUMBER OF SEQUENCES: 23
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: TX
; COUNTRY: USA
; ZIP: 77210-4433
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/505,307
; FILING DATE: 24-NOV-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/017,114
; FILING DATE: 12-FEB-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hodgins, Daniel S.
; REGISTRATION NUMBER: 31,026
; REFERENCE/DOCKET NUMBER: UTSK:231
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (713) 789-2679
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 193 amino acids
; TYPE: amino acid
; STRANDEDNESS:
; TOPOLOGY: linear
;
US-08-505-307-19

```

```

Query Match      10.28; Score 57; DB 3; Length 193;
Best Local Similarity 23.28; Pred. No. 11;
Matches 19; Conservative 13; Mismatches 12; Indels 38; Gaps 3;

QY 31 KEAQQITIEQTIKGNKKEISRFSGSADSIISFMIVYIKFGHTAILAPNRMQETLSLI 90
Db 4 OEODQOLVE-----RVQRGDKRAFDLVLTQTH-----KILGLIV 38
QY 91 SFL-----WVKPYR 99
|: :| :| :|
Db 39 RFVHDAQEADVDVAQEAFAFIKAYR 60

```

Search completed: July 30, 2002, 15:38:55
Job time: 184 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:37:11 ; Search time 16.29 seconds
(Without alignments)
654.752 Million cell updates/sec

Title: US-10-048-197-2

Perfect score: 558
Sequence: 1 MLNHHIRLTISALLTALLVT.....FLWVKPRPKNLSPYLTKAKA 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117.5	21.1	131	2 B91067	hypothetical lipop
2	78	14.0	289	2 H95216	cation efflux syst
3	78	14.0	299	2 F98080	hypothetical prote
4	74	13.3	328	2 AH2128	iron(III) diclrat
5	73	13.1	112	2 B81969	hypothetical prote
6	72.5	13.0	138	2 S74868	hypothetical prote
7	71.5	12.8	226	2 C70078	hypothetical prote
8	71	12.7	568	2 T49962	hypothetical prote
9	70.5	12.6	216	2 T40178	60s ribosomal prot
10	70	12.5	518	2 E71653	protein-export mem
11	69.5	12.5	314	1 G64145	hypothetical prote
12	69.5	12.5	337	2 T02532	hypothetical prote
13	69	12.4	530	2 T29418	hypothetical prote
14	68.5	12.3	391	2 C71498	probable gen. secr
15	68.5	12.3	405	2 E71179	hypothetical prote
16	68	12.2	430	2 E86873	proteinase (import
17	67.5	12.1	134	2 B64062	opacity-associated
18	67.5	12.1	210	2 E83816	late competence op
19	67.5	12.1	216	2 T40848	60s ribosomal prot
20	67	12.0	455	1 T21089	acid phosphatase (
21	67	12.0	488	1 H64313	corrinoid/iron-sul
22	67	12.0	799	1 B86735	hypothetical prote
23	67	12.0	819	2 G69801	hypothetical prote
24	66.5	11.9	92	2 B85911	hypothetical prote
25	66.5	11.9	310	2 G97345	oligopeptide ABC t
26	66.5	11.9	364	2 S48121	transcription fact
27	66.5	11.9	364	2 T50007	transcription fact
28	66.5	11.9	429	2 G71937	hypothetical prote
29	66	11.8	386	2 T01049	hypothetical prote

30	66	11.8	458	2 F71698	hypothetical prote
31	66	11.8	599	2 E36792	hypothetical prote
32	66	11.8	1828	2 B59254	myosin heavy chain
33	66	11.8	1855	2 A59254	myosin heavy chain
34	65.5	11.7	103	2 T58936	hypothetical prote
35	65.5	11.7	161	2 T08739	hypothetical prote
36	65.5	11.7	36	2 G69305	hypothetical prote
37	65.5	11.7	319	2 S62196	hypothetical prote
38	65.5	11.7	345	2 F90239	anthranilate phosp
39	65.5	11.7	419	2 B95106	conserved hypotet
40	65.5	11.7	425	2 H82183	conserved hypotet
41	65.5	11.7	629	2 T18227	hypothetical prote
42	65.5	11.7	690	2 H71237	hypothetical prote
43	64.5	11.6	196	2 G72423	hypothetical prote
44	64.5	11.6	419	2 D97974	conserved hypotet
45	64.5	11.6	492	2 T06875	preprotein translo

ALIGNMENTS

```

RESULT 1
B91067
hypothetical lipoprotein [imported] - Escherichia coli (strain O157:H7, substrain R1M
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 18-Jul-2001
C:Accession: B91067
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kuwakawa, K.; Ishii, K.; Yokoyama, K.; Han, C
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and g
A:Reference number: A96629; MUID:21156231; PMID:11258796
A:Accession: B91067
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-131 <HAY>
A:Cross-references: GB:BA000007; PIDN:BA836929.1; PID:G13362977; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain R1M 0509952
C:Genetics:
A:Gene: ECs3506

Query Match 21.1%; Score 117.5; DB 2; Length 131;
Best Local Similarity 29.7%; Pred. No. 4.3e-05;
Matches 27; Conservative 19; Mismatches 30; Indels 15; Gaps 1;

Oy 17 LVTGCVSTGVAMKEONQOTIEOTIIKGTNKQDISRFGSADSISFMIVIRGHT-- 74
      | : | | | | | : | : | : | : | : | : | : | : | : | : | : |
Db 19 LTLGCSVGNQSIKNETOESVKTIVKGTAKODVLASFGEPSRSILIDEEOMSYTMV 78
Oy 75 -----AIIAPNRQGLSLISF 92
      | | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db 79 NSQSKATSFIPVGLAGGADSRKSLTVSF 109

RESULT 2
H95216
cation efflux system protein [imported] - Streptococcus pneumoniae (strain TIGR4)
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: H95216
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; H
on, J.D.; Umeyam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapp
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morris
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: H95216
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-299 <KUR>
A:Cross-references: GB:AE005672; PIDN:AAK75929.1; PID:G14973359; GSPDB:GN00164; TIGR:

```

A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP1857

Query Match	14.08;	Score 78;	DB 2;	Length 299;
Best Local Similarity	28.38;	Pred. No. 1.5;		
Matches 36;	Conservative 18;	Mismatches 39;	Indels 34;	Gaps 6;

```
QY      10 ISALLTA-LVTVGCYST--GNVMKKEEDNOTIOTII-----RGKTN 46  
       ||::|||::|||::|||::|||::|||::|||::|||::|||  
Db      78 LGALVTAVTILVTGTVLEIVLEWTKILHPQRYNDEGIIMLGITINLTNLASLVVGKKR 137  
  
QY      49 KOEISSREGSADSISE----MIVIKFGHTAILAPNRQOELISLTSFLMWKPYPRPKNL 103  
       |::|::|::|::|::|::|::|::|::|::|::|::|:  
Db      138 NESITSLFLEDITIGWAIVIIMAIIVALTFTDVIIDP-----LSLVSIFSLKAUR- 190  
  
QY      104 SFYLTFAK 110  
       |::|  
Db      191 -FWSFLK 196
```

RESULT	3
1000000	

hypothetical protein czcd [imported] - Streptococcus pneumoniae (strain R6)
F98080
C:Species: Streptococcus pneumoniae
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 22-Oct-2001
C:Accession: F98080
R:Roskams, J.A.; Albourn Jr., W.; Arnold, J.; Blaszcak, L.; Buregelt, S.; Dehoff, B.S.; E
y, R.; Leblanc, D.J.; Lee, L.N.; Leikowitz, E.J.; Lu, J.; Matsushima, P.; McAhren, S.; M
y, P.; Sun, P.M.; Winkler, M.E.
J. Bacteriol. 183, 5709-5717, 2001
A:Authors: Yang, Y.; Young-Bellido, M.; Zhao, G.; Zook, C.; Baltz, R.H.; Jaskunas, S.R.;
A:Title: Genome of the Bacterium Streptococcus pneumoniae Strain R6.
A:Reference number: A97872; MUID:21429245; PMID:11544254
A:Accession: F98080
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-299 <KUD>
A:Cross-references: GB:AB007317; PID:AAL00475.1; PID:g15459345; GSPDB:GN00174
A:Genetics: czcd
A:Gene: czcd

```

Query Match          14.0%; Score 78; DB 2; Length 299;
Best local Similarity 28.3%; Pred. No. 1.5;
Matches 36; Conservative 18; Mismatches 39; Indels 34; Gaps 6;

QY 10 ISALLTA-LVATGCVST--GNVAMKEONOQTIEQITII-----KGKN 48
      :||:||||| | | | | | | | | | | | | | | | | | | | |
Db 78 LGALVTAVIIVTSGVIVILENFKILHPQVNDGILMLGIATITNLLASLVGKRTK 137
      :||:||||| | | | | | | | | | | | | | | | | | | | |

QY 49 KOEISSRFSGSDSISF-----MIVIKFGHAILAPRQWEIISLISFLVAPRPNKL 103
      : | | | | | | | | | | | | | | | | | | | | | | | |
      : | | | | | | | | | | | | | | | | | | | | | | | |
Db 138 NESIISLHFLEDITLGVAVIIMATVLRPTDWTILDp-----LLSLVISFILSKALPR-- 190
      : | | | | | | | | | | | | | | | | | | | | | | | |

QY 104 SEYLTAK 110
      ||: || |
Db 191 -FWSTLK 196

RESULT 4
AH2128
Iron(III) dicitrate-binding periplasmic protein [imported] - Anabaena sp. (strain PCC 7120)
C:Species: Anabaena sp.
A:Note: Anabaena sp. (Strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AH2128
R:Kaneoko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, S.;
Hakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium Anabaena
A:Reference number: AB1807; MUID:21595285; PMID:11759840

```

A:Accession: AH2128
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1,328 <KIR>
A:Cross-references: GB:BA000019; PIDN:BA674282.1; PID:917131675; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2583
C:Superfamily: ferrichrome-iron transport protein feoB

Query Match	13.3%	Score 74:	DB 2:	Length 328:
Best Local Similarity	31.4%	Pred. No. 4.3:		
Matches	22:	Conservative	14:	Mismatches 26: Indels 8: Gaps 3:
Oy	1	MLNNH-IRLNHTSALLTALLVTGC-VSTGNVANKRQNOOTIQTIIKKTGNKQELSR---	55	
Db	18	ILNNHQLKILFSLITALLIVIGCEMSTPNVNTINSVNTSEMRYVKHTMGETKILPRLP	77	
Oy	56	--FGSADSI	62	
Db	78	VVVLGGLDNI	87	

RESULT 5

B81969
 Hypoetical protein NMA0510 [imported] - Neisseria meningitidis (strain Z2491 serog
 C:Species: Neisseria meningitidis
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
 C:Accession: B81969
 R:Parikhall, I.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo
 R:Parikhall, S.; Jørgensen, K.; Leather, S.; Moulé, S.; Mungall, K.; Quail, M.A.; Rajander
 Nature 404, 502-505, 2000
 A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis Z2491
 A:Reference number: AB1775; MIMD:20222556
 A:Accession: B81969
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-112 <PAR>
 A:Cross-references: GB:AL162753; GB:AL157959, NID:g7379120; PIDN:CAB83802.1; PID:g737
 A:Experimental source: serogroup A, strain Z2491
 C:Genetics:
 A:Gene: NMA0510

	Query Match	13.1%	Score 73;	DB 2;	Length 112;	
	Best Local Similarity	31.2%;	Pred. No. 1.7;			
	Matches	20;	Conservative	9;	Mismatches	35;
					Indels	0;
					Gaps	0;
OY	7	RUTSALTLTAVTGCVSTGNVAMKEQNOQTIECTIIRKTKNKOELSSRGASDSIFMI	66			
		: :	: :	: :	: :	: :
Db	4	RIIVSAFVALALAGCGSINNTVSQKIQEAPAAVLGSPNAVKISINRSNEGIRINFPA	63			
OY	67	VVIK 70				
Db	64	TWCK 67				

RESULT 6

S74868
hypothetical protein s11158 - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74868
R:Ranko, T., Sato, S., Kotani, H., Tanaka, A., Asamizu, E., Nakamura, Y., Miyajima, O., K., Okumura, S., Shimo, S., Takeuchi, C., Wada, T., Watanabe, A., Yamada, M., Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S74868
A:Status: preliminary

A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: secD; RP586
 C:Superfamily: protein export membrane protein secD

Query Match 12.5%; Score 70; DB 2; Length 518;
 Best Local Similarity 26.5%; Pred. No. 19;
 Matches 31; Conservative 21; Mismatches 39; Indels 26; Gaps 6;
 Oy 8 LTTLSALTALVTCVSTGVNVMKEONQOTIEOTIKGKTKNKOESSRFGSA-----DS 61
 Db 404 LTPGNGIITLTGMADVANDALVYER-----IKRELKGVSNLVAITGFSAPATIDSN 459
 Oy 62 ISFMIVIVIK---FGHTAI-----IAPNMQEIL--SLTISLWKPYPKPKL 103
 Db 460 ITTILVAFALYIFGVGAIKGFVAVALFTIGLISMSFSAITIKLLID-VWVKYKPKKL 515

RESULT 11

G64145
 hypothetical protein H10230 - Haemophilus influenzae (strain Rd KM20)
 C:Species: Haemophilus influenzae
 C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
 C:Accession: G64145
 R:Flaeschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, J.
 C:Flaeschmann, R.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.
 D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.
 Science 269, 496-512, 1995
 A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter,
 A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
 A:Reference number: A64000; MUID:95350630
 A:Accession: G64145
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Accession: G64145
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-314 <TIGR>
 A:Cross-references: GB:U02709; GB:I42023; NID:g1573190; PIDN:AAC21899.1; PID:g1573194;
 A:Note: best homolog was a hypothetical protein from Escherichia coli
 C:Superfamily: hypothetical protein H10230; tetratricopeptide repeat homology
 F:74-107/Domain: tetratricopeptide repeat homology <TT1>
 F:108-141/Domain: tetratricopeptide repeat homology <TT2>
 F:142-175/Domain: tetratricopeptide repeat homology <TT3>

Query Match 12.5%; Score 69.5; DB 1; Length 314;
 Best Local Similarity 27.1%; Pred. No. 12;
 Matches 19; Conservative 13; Mismatches 21; Indels 17; Gaps 2;
 Oy 2 LNHRIITISALLTALVTCG-----VSTGVNVMKEONQOT-----IEOTIK 44
 Db 6 LSRHFIVYLSLCAILLACVOSRGFGVSKNHNVLAEONPNTHFQEVMIYRLSQLLV 65
 Oy 45 GKTNKOESS 54
 Db 66 GKMSNEERAS 75

RESULT 12

T02532
 hypothetical protein At2g37660 [Imported] - Arabidopsis thaliana
 N:Alternate names: hypothetical protein F13M22.16
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 05-Mar-1999 #sequence_revision 05-Mar-1999 #text_change 16-Feb-2001
 C:Accession: T02532; E84795
 R:Rounsley, S.D.; Kaul, S.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes,
 submitted to the EMBL Data Library, June 1998
 A:Description: Arabidopsis thaliana chromosome II BAC F13M22 genomic sequence.
 A:Reference number: Z14677
 A:Accession: T02532
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-337 <ROU>
 A:Cross-references: EMBL:AC004684; NID:g3236234; PID:g3236248

A:Experimental source: columbia Columbia
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shema, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
 M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanacker, S.E.; Umayam, L.; Tallon,
 euss, D.; Nieman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
 Nature 402, 761-768, 1999
 A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
 A:Reference number: A84420; MUID:20083487

Query Match 12.5%; Score 69.5; DB 2; Length 337;
 Best Local Similarity 23.2%; Pred. No. 13;
 Matches 36; Conservative 26; Mismatches 38; Indels 55; Gaps 8;
 Oy 3 NHHIRLTISALT-----ALVTGCVS--TGNYA---MKRONQOTIEOTIKGKTKNKOESS 54
 Db 61 NRRVSVTVSAATTEPLEVLTGTAGGRTGQIVYKKLREBQFVARBLVTKESKEKING 120
 Oy 55 RGSADISFMIVIVIKFGHTAILAPNMQEILSLTISFLWKPYP----- 100
 Db 121 E-----VFQIDINDTASINP-AVEGIDALVILTSAPVPMKPGFDPKSGRPEFF 171
 Oy 101 -----KN-----ISFYLTAKA 111
 Db 172 DGAAPPEQWIGORKNOIDAGVDDGLSFVYAKA 206

RESULT 13

T29418
 hypothetical protein F21F8.11 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
 C:Accession: T29418
 R:Wilson, R.; Favello, A.; Le, T.T.
 submitted to the EMBL Data Library, April 1997
 A:Description: The sequence of C. elegans cosmid F21F8.
 A:Reference number: Z20618
 A:Accession: T29418
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-530 <MIL>
 A:Cross-references: EMBL:U97000; PIDN:AAC47997.1; GSPDB:GN00023; CESP:F21F8.11
 A:Experimental source: strain Bristol N2; clone F21F8
 C:Genetics:
 A:Gene: CESP:F21F8.11
 A:Map position: 5
 A:Introns: 16/1; 64/2; 112/3; 344/3; 417/1; 452/1

Query Match 12.4%; Score 69; DB 2; Length 530;
 Best Local Similarity 21.7%; Pred. No. 24;
 Matches 30; Conservative 25; Mismatches 47; Indels 36; Gaps 5;
 Oy 5 HIRLTISALLTALVTCVSTGVNVMKEONQOTIEOTIKGKTKNKOESSRFGSAD----- 60
 Db 44 HIGLSLSCMCNSTAV-ALMNTNNATLVGISTILMSLENKISE---SQELGGDDGES 99
 Oy 61 -----SISFMIVIVIK-----FGHTAILAPNMQEILSLTISFLWKPYP----- 100
 Db 100 CTKLESKVIXDYGGFINSVSMOGIYSAALFGIFFSYPAGVLDVRSARHILSVAILM 159
 Oy 93 LWWKPYRPKNLSFYLTAK 110
 Db 160 LTIASLMPVLISVIYGEK 177

RESULT 14

C71499

probable gen. secretion protein F - Chlamydia trachomatis (serotype D, strain UW3/Cx)

C:Species: Chlamydia trachomatis

C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 17-Mar-2000

C:Accession: C71499

R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, Science 282, 754-759, 1998

A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis

A:Reference number: A71570; MUID:9900809

A:Accession: C71499

A>Status: Preliminary

A:Molecule type: DNA

A:Residues: 1-391 <ARN>

A:Cross-references: GB:AE00327; GB:AE00273; NID:93328999; PIDN:BA68172.1; PID:9332901

A:Experimental source: serotype D, strain UW-3/Cx

C:Genetics:

A:Gene: g9pF

C:Superfamily: secretion protein xcps

Query Match

12.3%; Score 68.5; DB 2; Length 391;

Best Local Similarity 21.7%; Pred. No. 20;

Matches 20; Conservative 22; Mismatches 39; Indels 11; Gaps 3;

OY 14 LTALLVGCSTGNVAMKEONQOTIEQTIKGTNKOEISSRFGSADISFMI----- 66

DB 275 LIEALVTLGCEAVSODELREELQEVY-QAVVRGSGLSRELSHRTWTPKLVIGWALGEESG 333

OY 67 -VVIKFGHTAILAPNRMOEILSLISFLWKP 97

DB 334 DLAVFAHVAQIYVEDIORVLTWTA--WCQP 363

RESULT 15

E71179

hypothetical protein PH1714 - Pyrococcus horikoshii

C:Species: Pyrococcus horikoshii

C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000

C:Accession: E71179

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin M.; Ohtuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon: Pyrococcus horikoshii

A:Reference number: A71000; MUID:98344137

A:Accession: E71179

A>Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-405 <KAM>

A:Cross-references: GB:AP000007; NID:93236134; PIDN:BA30828.1; PID:93258145

A:Experimental source: strain OT3

C:Genetics:

A:Gene: PH1714

Query Match

12.3%; Score 68.5; DB 2; Length 405;

Best Local Similarity 22.5%; Pred. No. 20;

Matches 27; Conservative 17; Mismatches 33; Indels 43; Gaps 4;

OY 6 IRLTISALTLVLTGCVSTGNVAMKEONQOTIEQTIKGTNKOEISSRFGSADISF- 64

DB 5 LALSTIGVLAISVAGCTGGG-----TOTOTOGKSIVAILFDVGGRGDLSFN 54

OY 65 -----MIVVIFGHTAILAPNRMOEILSL-----IISFLWKP 97

DB 55 DMAYLGARAKKELGVKIEY-----MTPKSKEDVPLLEQLAKSKEYDILVLVGFLLWTPP 109

Search completed: July 30, 2002, 15:39:18

Job time: 127 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:38:56 ; Search time 11.9 seconds
(without alignments) 361.166 Million cell updates/sec

Title: US-10-048-197-2
Perfect score: 558
Sequence: 1 MLNHHIRLTSLTALTALVT.....FLWVKPYRPNLSFYLTAKA 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	71.5	12.8	226 1	YXIP_BACSU
2	70.5	12.6	216 1	R10A_SCHPO
3	70	12.5	518 1	SECD_RICPR
4	69.5	12.5	314 1	NLPI_HAEIN
5	69.5	12.5	337 1	Y230_ARATH
6	68.5	12.3	405 1	VH14_PYRHO
7	67.5	12.1	134 1	OAPB_HAEIN
8	67.5	12.1	216 1	R10B_SCHPO
9	66	11.8	370 1	RF1_MYCCA
10	66	11.8	599 1	VG58_HSV11
11	66	11.8	1855 1	MY5A_HUMAN
12	65.5	11.7	345 1	TRPD_SUISO
13	64.5	11.6	492 1	SECT_CYPAP
14	64	11.5	157 1	Y406_MYCPN
15	64	11.5	880 1	RPAL_SUISO
16	64	11.5	1828 1	MY5A_RAT
17	64	11.5	1829 1	MY5A_CHICK
18	64	11.5	1853 1	MY5A_MOUSE
19	63.5	11.4	986 1	DROL_NPYBM
20	63	11.3	254 1	T2H1_HAEPA
21	63	11.3	262 1	VF10_HAEIN
22	62.5	11.2	984. 1	DROL_NPVAC
23	62.5	11.2	1145 1	PR22_YEAST
24	62.5	11.2	3206 1	POLG_PSBMV
25	62	11.1	433 1	ORC4_MOUSE
26	62	11.1	442 1	YMD8_YEAST
27	62	11.1	1062 1	YAF1_YEAST
28	61.5	11.0	467 1	LOLS_DROME
29	61.5	11.0	894 1	LOLS_DROME
30	61	10.9	265 1	SYNP_RAT
31	61	10.9	288 1	UBIA_PROST
32	61	10.9	1940 1	MYH3_HUMAN
33	61	10.9	1940 1	MYH3_RAT

34	60.5	10.8	201 1	OSMY_ECOLI
35	60	10.8	335 1	FLIG_THEMA
36	60	10.8	359 1	MCAL_CRIGR
37	60	10.8	367 1	YC36_HAEIN
38	60	10.8	475 1	NCAP_STNV
39	60	10.8	571 1	TACY_STRPY
40	60	10.8	603 1	NU5M_RABIT
41	60	10.8	654 1	RPC3_YEAST
42	60	10.8	1091 1	XYI_TREPA
43	59.5	10.7	330 1	EXL2_HUMAN
44	59.5	10.7	414 1	YBBC_BACSU
45	59.5	10.7	568 1	MET3_ASPTE

ALIGNMENTS

RESULT 1	YXIP_BACSU	STANDARD;	PRT;	226 AA.
AC	P42307:			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1995 (Rel. 32, Last sequence update)			
DT	01-MAR-2002 (Rel. 41, Last annotation update)			
DE	Hypothetical protein yxip precursor.			
GN	YXIP OR S3B.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1423;			
NP	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-168 / BGSC1A1;			
RX	MEDLINE-97124196; PubMed-8969509;			
RA	Yoshida K.-I., Shindo K., Sano H., Seki S., Fujimura M., Yanai N.,			
RT	Miwa Y., Fujita Y.;			
RT	"Sequencing of a 65 kb region of the Bacillus subtilis genome			
RT	containing the ltc and cel loci, and creation of a 177 kb contig			
RT	covering the gnt-sacXY region."			
RL	Microbiology 142:3113-3123(1996).			
RN	[2]			
RP	SEQUENCE OF 176-226 FROM N.A.			
RC	STRAIN-BR151;			
RA	Stuelse J., Schmetz K., Krieg M., Krueger S., Hecker M., Rak B.;			
RL	Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; D83026; BAAL1695.1; -			
DR	EMBL; 228340; -; NOT_ANNOTATED_CDS.			
DR	EMBL; 299124; CAB15945.1; -			
DR	Subtilist; BC11145; yxip.			
DR	PROSITE; PS00013; PROKAR_LIPOPROTEIN; UNKNOWN_1.			
KW	Hypothetical protein; Signal; Complete proteome.			
FT	SIGNAL 1 18			
FT	CHAIN 19 226			
FT	SEQUENCE 226 AA; 25697 MW; 49D7A86DF4400D08 CMC64;			

Query Match 12.8%; Score 71.5; DB 1; Length 226;
Best Local Similarity 36.2%; Pred. No. 1.9;
Matches 17; Conservative 13; Mismatches 16; Indels 1; Gaps 1;

Qy 6 IRLTSLTALTYGCVSTGVNVA-KRONOQTITQTIKGTNKE 51
DB 4 IGLTSLTALTYVMSACESEGEAOMFADCDKQTVAKPSSKK 50

RESULT 2
 ID R10A_SCHPO STANDARD: PRT: 216 AA.
 AC 014363;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE 60S ribosomal protein L1-A (L10A).
 GN RPLA OR SPBC30D10.18C.
 OS Schizosaccharomyces pombe (fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomyces.
 OX NCBI_Taxid=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Wood V., Rajandream M.A., Barrell B.G., Lauber J., Hiltbert H.,
 RA Duesterhoeft A.; to the EMBL/GenBank/DBJ databases.
 RL Submitted (JUL-1997) to the EMBL/GenBank/DBJ databases.
 CC -1- MISCELLANEOUS: THERE ARE TWO GENES FOR L1 IN S. POMBE.
 CC -1- SIMILARITY: BELONGS TO THE L1P FAMILY OF RIBOSOMAL PROTEINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: 297992; CAB10813.1;
 DR InterPro: IPR002143; Ribosomal_L1.
 DR Pfam: PF00687; Ribosomal_L1: 1.
 DR ProDom: PD001314; Ribosomal_L1: 2.
 DR PROSITE: PS01199; RIBOSOMAL_L1: 1.
 KW Ribosomal protein; Multigene family.
 SQ SEQUENCE 216 AA; 23884 MW; 676EA3710A497165 CRC64;

Query Match 12.6%; Score 70.5; DB 1; Length 216;
 Best Local Similarity 26.6%; Pred. No. 2.4;
 Matches 25; Conservative 16; Mismatches 24; Indels 29; Gaps 4;

OY 6 IRLTISALTLALVTGCVSTGVNAAKEQNOQTIEQTIKTKNKQKISSRFSADSIFM 65
 DB 150 VKSTIKFQLKVLCLG-VAVGHVDMAB-----EQLANLSLA--INFL 189
 OY 66 IVIKFEGHTAIAPNRMOELSLIISFLWKKPYR 99
 DB 190 VSLKKG-----WONIGSLVYKSTMGKPR 214

RESULT 3
 ID SECD_RICPR STANDARD: PRT: 518 AA.
 AC 092CWB;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Protein-export membrane protein secD.
 GN SECD OR RP586.
 OS Rickettsia prowazekii.
 OC Bacteria; Proteobacteria; alpha subdivision; Rickettsiales;
 OC Rickettsiaceae; Rickettsia.
 OX NCBI_Taxid=782;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=MADRID E;
 RX MEDLINE=99039499; PubMed=9823893;
 RA Andersson S.G.E., Zomorodipour A., Andersson J.O.,
 RA Sichenitz-Ponten T., Alsmark U.C.M., Podowski R.M., Naeslund A.K.,

RA Eriksson A.-S., Winkler H.H., Kurland C.G.;
 RT "The genome sequence of Rickettsia prowazekii and the origin of
 RT mitochondria."
 RL Nature 396:133-140(1998).
 CC -1- FUNCTION: INVOLVED IN PROTEIN EXPORT.
 CC -1- SUBUNIT: PART OF THE PROKARYOTIC PROTEIN TRANSLLOCATION APPARATUS
 CC WHICH COMPRISE SECA, SECB, SECD, SECE, SECF, SECG AND SECH
 CC (BY SIMILARITY)
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (by similarity).
 CC -1- SIMILARITY: BELONGS TO THE SECD/SECF FAMILY. SECD FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AJ235272; CA15031.1;
 DR InterPro: IPR003335; SecD_SecF.
 DR Pfam: PF02355; SecD_SecF. 1.
 DR Protein transport; Translocation; Transmembrane; Membrane;
 KW Complete proteome. 29
 FT TRANSMEM 9 POTENTIAL.
 FT TRANSMEM 356 376 POTENTIAL.
 FT TRANSMEM 377 397 POTENTIAL.
 FT TRANSMEM 406 426 POTENTIAL.
 FT TRANSMEM 464 484 POTENTIAL.
 FT TRANSMEM 486 506 POTENTIAL.
 SQ SEQUENCE 518 AA; 56712 MW; 86D6FC2A5B5DE2AA CRC64;

Query Match 12.5%; Score 70; DB 1; Length 518;
 Best Local Similarity 26.5%; Pred. No. 6.9;
 Matches 31; Conservative 21; Mismatches 39; Indels 26; Gaps 6;

OY 8 LRTISALTLALVTGCVSTGVNAAKEQNOQTIEQTIKTKNKQKISSRFSGA-----DS 61
 DB 404 LTLPGTIGLITLGMADVANDVLIYER-----IKEELHKGVSNLYARTGFESAFATITDSN 459
 OY 62 ISFMIVIK---FGHTAI-----LAPNRMOEL--SLIISFLWKKPYRPNL 103
 DB 460 ITTLIVAFALYIRGVGAIKGFVALRTIGLITSSMFSATITTKLLID-VWYKPKPKKL 515

RESULT 4
 ID NLP1_HAEIN STANDARD: PRT: 314 AA.
 AC P44585;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Lipoprotein nlp1 homolog precursor.
 GN NLP1 OR HI0230.
 OS Haemophilus influenzae.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Haemophilus.
 OX NCBI_Taxid=727;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=RD / KW20 / ATCC 51907;
 RX MEDLINE=95350630; PubMed=7542800;
 RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
 RA Kienle A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M.,
 RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
 RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
 RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
 RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
 RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
 RA Venter J.C.;
 RT "Whole-genome random sequencing and assembly of Haemophilus

```

RT Influenzae Rd.*
RL Science 269:496-512(1995).
CC -1- FUNCTION: MAY BE INVOLVED IN CELL DIVISION (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (By similarity).
CC -1- SIMILARITY: CONTAINS 3 TPR REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U33709; AAC21899.1; -.
DR TIGR: H10230; -.
DR InterPro: IPR001440; TPR.
DR Pfam: PF00515; TPR; 3.
KW Cell division; Membrane; Lipoprotein; Repeat; TPR repeat; Signal;
KM Complete proteome.
FT SIGNAL 1 25 BY SIMILARITY.
FT CHAIN 26 314 LIPOPROTEIN NLPI HOMOLOG.
FT LIPID 26 26 N-ACYL DIGLYCERIDE.
FT REPEAT 74 107 TPR 1.
FT REPEAT 108 141 TPR 2.
FT REPEAT 142 175 TPR 3.
SQ SEQUENCE 314 AA; 3643 MW; 0B87CAE286CC5722 CRC64;

```

```

Query Match 12.5%; Score 69.5; DB 1; Length 314;
Best Local Similarity 27.1%; Pred. No. 4.5;
Matches 19; Conservative 13; Mismatches 21; Indels 17; Gaps 2;

```

```

QY 2 LNHHIRLTISALLTALVTGC-----VSTGNVAMKEQNOOT-----IEQRTIK 44
   1 : : : : : 11 : : : : : 1 : : : : : 1 : : : : :
DB 6 LSRHFIVYLFSCAILLACGCVSRGKGFVSKNHYVLAENQNTHEQDEMTIVRLSQVLLV 65
   1 : : : : : 1 : : : : : 1 : : : : : 1 : : : : :
QY 45 GKTNKOEISS 54
   1 : : : : : 1 : : : : : 1 : : : : : 1 : : : : :
DB 66 GKMSNEERAS 75

```

```

RESULT 5
Y230_ARATH STANDARD: PRT; 337 AA.
ID Y230_ARATH
AC 080934;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Protein At2g37520, chloroplast precursor.
GN AT2G37520 OR F13M22.16.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosoids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=20083487; PubMed=10617197;
RA Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA Motilal K.S., Cronin L.A., Shen M., Pal G., Van Aken S., Umayam L.,
RA Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA Norman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA Mergam W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA Venter J.C.;
RT "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT thaliana."
RL Nature 402:761-768(1999).
CC -1- SUBCELLULAR LOCATION: Chloroplast (Potential).

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AC004684; AAC23636.1; -.
DR SWISS-2DPAGE: 080934; ARATH.
KW Chloroplast; Transist peptide.
FT TRANSIT 1 337 CHLOROPLAST (POTENTIAL).
FT CHAIN ? 337 PROTEIN AT2G37520.
SQ SEQUENCE 337 AA; 36085 MW; 748838ECB4BA07CD CRC64;

```

```

Query Match 12.5%; Score 69.5; DB 1; Length 337;
Best Local Similarity 23.2%; Pred. No. 4.9;
Matches 36; Conservative 26; Mismatches 38; Indels 55; Gaps 8;

```

```

QY 3 NHHIRLTISALLT-----ALVTGCVS-TGNVA---MKEQNOOTIEQRTIKTKNKOEISS 54
   1 : : : : : 1 : : : : : 1 : : : : : 1 : : : : :
DB 61 NRRSVTVSAATTEPLVLTGAGGRTGQIVYKLRKRSQFVARKGLVRTKSEKING 120
   1 : : : : : 1 : : : : : 1 : : : : : 1 : : : : :
QY 55 RFGSADSIEMIVIKFGHTAILAPNRMOELISLISFLWKPYRP----- 100
   1 : : : : : 1 : : : : : 1 : : : : : 1 : : : : :
DB 121 E-----DF-----VFIGDIRDTASINP-AVEGIDALVILTSANVPQMKFGDPKSGARPEFF 171
   1 : : : : : 1 : : : : : 1 : : : : : 1 : : : : :
QY 101 -----KN-----LSFYLAKA 111
   1 : : : : : 1 : : : : : 1 : : : : : 1 : : : : :
DB 172 DDGAVPEQVDWIGKKNQIDAGVDDGSLFVYTAKA 206

```

```

RESULT 6
YH14_PYRHO STANDARD: PRT; 405 AA.
ID YH14_PYRHO
AC 059403;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical Lipoprotein PH1714 precursor.
GN PH1714 OR PHAM037.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3;
RX MEDLINE=98344137; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Haino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohnuki Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Maechli Y., Shizuwa H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
RL DNA Res. 5:55-76(1998).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a lipid anchor
CC (Potential).
CC -1- SIMILARITY: BELONGS TO THE BMP FAMILY OF LIPOPROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AP000007; BAA30828.1; -.
DR InterPro: IPR003760; Bmp.

```

DR TIGR; HI0331; -.

5
F
C
C
E
I
A

5
F
C
C
E
I
A

```

DE Peptide chain release factor 1 (RF-1).
GN PRA.
OS Mycoplasma capricolum.
OC Bacteria: Firmicutes; Bacillus/Clostridium group; Mollicutes;
CC Entomoplasmataceae.
OX NCBI_TaxID=2095;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 27343 / KID:
RX MEDLINE=96166911; PubMed=8635731;
RA Ingaki Y., Bessho Y., Hori H., Osawa S.;
RT "Cloning of the Mycoplasma capricolum gene encoding peptide-chain
  release factor."
RL Gene 169:101-103(1996).
CC -1- FUNCTION: PEPTIDE CHAIN RELEASE FACTOR 1 DIRECTS THE TERMINATION
  OF TRANSLATION IN RESPONSE TO THE PEPTIDE CHAIN TERMINATION
CC CODONS UAG AND UAA (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE PROKARYOTIC AND MITOCHONDRIAL RELEASE
  FACTORS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
  or send an email to license@isb.slb.ch).
CC -----
DR EMBL; D50100; BAA08795.1;
DR InterPro: IPR000352; Pep_rel_factor_1.
DR Pfam: PF00472; RF-1; 1.
DR PROSITE: PS00745; RF_PROK_I; 1.
DR Protein biosynthesis.
SQ SEQUENCE 370 AA; 42186 MW; 7438BACAD6ED20E4 CRC64;

Query Match 11.8%; Score 66; DB 1; Length 370;
Best Local Similarity 26.3%; Pred. No. 12;
Matches 15; Conservative 12; Mismatches 30; Indels 0; Gaps 0;

OY 5 HIRLTALLTALLVTGCVSTGNVAMKEONQOTIEQTIKTKNKOEISSRFGSADS 61
Db 238 HVMNTDSAVRITHLPTGIVTSDQRSQHDNKMIDAMTFLRAKVEAEVEKQQAQADA 294

RESULT 10
VG58_HSV11 STANDARD; PRT; 599 AA.
AC 000157;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 01-DEC-1992 (Rel. 24, Last annotation update)
DE Hypothetical gene 58 protein.
GN 58.
OS Ictalurid herpesvirus 1 (Channel catfish virus) (CCV).
OC Viruses: dsDNA viruses, no RNA stage; Herpesviridae;
CC Unclassified Herpesviridae.
OX NCBI_TaxID=10401;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=AUBURN 1;
RX MEDLINE=92087490; PubMed=1727613;
RA Davison A.J.;
RT "Channel catfish virus: a new type of herpesvirus.";
RL Virology 186:9-14(1992).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
  or send an email to license@isb.slb.ch).

```

```

CC or send an email to license@isb.slb.ch).
CC -----
DR EMBL; M75136; AAA88161.1;
DR PIR; E36792; E36792.
KW Hypothetical protein.
SQ SEQUENCE 599 AA; 67476 MW; 991C2B524F0F6BEF CRC64;

Query Match 11.8%; Score 66; DB 1; Length 599;
Best Local Similarity 33.9%; Pred. No. 21;
Matches 21; Conservative 9; Mismatches 24; Indels 8; Gaps 2;

OY 32 EONQOTIEQTIKTKNKOEISSRFGSADSI--FMIV-----IKRGHTAILAPNRQ 83
Db 28 ELNNGEITNTIKIKRKGMTGFKSSRFGATESITNDPDLFNGALVKTGHLDAKPVTWA 87

OY 84 EI 85
Db 88 KL 89

RESULT 11
MY5A_HUMAN STANDARD; PRT; 1855 AA.
AC 09Y411; 09UE30; 09UE31; 007902; Q16249; O60653;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin Va (Myosin 5A) (Dilute myosin heavy chain, non-muscle)
DE (Myosin heavy chain 12) (Myosin).
OS MYO5A OR MYH12.
GN Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Skin;
RA Meurers B.H., Zimmermann R., Vosberg H.P.;
RT "The complete cDNA for human myosin heavy chain 12, a class V
  myosin."
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT CVS-1246.
RX MEDLINE=97351514; PubMed=9207796;
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
RT "Griscelli disease maps to chromosome 15q21 and is associated with
  mutations in the myosin-Va gene."
RL Nat. Genet. 16:289-292(1997).
RN [3]
RP ERRATUM.
RA Pastural E., Barrat F.J., Dufourcq-Lagelouse R., Certain S., Sanal O.,
RA Jabado N., Seger R., Griscelli C., Fischer A., de Saint Basile G.;
RL Nat. Genet. 23:373-373(1999).
RN [4]
RP SEQUENCE OF 638-1477 FROM N.A. (ISOFORM 2).
RC TISSUE=fetal brain;
RX MEDLINE=94245227; PubMed=8188282;
RA Engle L.J., Kennett R.H.;
RT "Cloning, analysis, and chromosomal localization of myosin (MYH12),
  the human homologue to the mouse dilute gene."
RL Genomics 19:407-416(1994).
RN [5]
RP SEQUENCE OF 1061-1498 FROM N.A. (ISOFORM 2).
RC TISSUE=Brain;
RX MEDLINE=95136715; PubMed=7835087;
RA Moore K.J., Testa J.R., Francke U., Milatovich A., Copeland N.G.,
RA Jenkins N.A.;
RT "Cloning and regional assignment of the human myosin heavy chain 12
  (MYH12) gene to chromosome band 15q21."
RL Cytogenet. Cell Genet. 69:53-58(1995).
RN [6]

```

RP SEQUENCE OF 1358-1460 FROM N.A. (ISOFORM 3).
 RA Edgar A.J., Bennett J.P.;
 RT "Inhibition of dendrite formation in melanocytes transiently
 RT transfected with antisense DNA to myosin V.";
 RT Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.
 RN [7]
 RP FUNCTION.
 RX MEDLINE=99376094; PubMed=10448864;
 RA Mehta A.D., Rock R.S., Rief M., Spudich J.A., Mooseker M.S.,
 RA Cheney R.E.;
 RT "Myosin-V is a processive actin-based motor.";
 RT Nature 400:590-593(1999).
 CC -1- FUNCTION: PROCESSIVE ACTIN-BASED MOTOR THAT CAN MOVE IN LARGE
 CC STEPS APPROXIMATING THE 36-NM PSEUDO-REPEAT OF THE ACTIN FILAMENT.
 CC MAY BE INVOLVED IN MELANOSOME TRANSPORT, OR ALTERNATIVELY, IT MAY
 CC BE REQUIRED FOR SOME POLARIZATION PROCESS INVOLVED IN DENDRITE
 CC FORMATION.
 CC -1- SUBUNIT: MAY BE A HOMODIMER, WHICH ASSOCIATES WITH MULTIPLE
 CC CALMODULIN OR MYOSIN LIGHT CHAINS.
 CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS; 1 (SHOWN HERE), 2 AND 3; ARE
 CC PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- DISEASE: DEFECTS IN MYO5A ARE A CAUSE OF GRISCELLI SYNDROME (GS).
 CC GS IS A RARE AUTOSOMAL RECESSIVE DISORDER THAT RESULTS IN
 CC PIGMENTARY DILUTION OF THE SKIN AND HAIR, THE PRESENCE OF LARGE
 CC CLUMPS OF PIGMENT IN HAIR SHAFTS, AND AN ACCUMULATION OF
 CC MELANOSOMES IN MELANOCYTES. MOST PATIENTS ALSO DEVELOP AN
 CC UNCONTROLLED T LYMPHOCYTE AND MACROPHAGE ACTIVATION SYNDROME,
 CC KNOWN AS HEMOPHAGOCYTIC SYNDROME, LEADING TO DEATH IN THE ABSENCE
 CC OF BONE MARROW TRANSPLANTATION.
 CC -1- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
 CC -1- SIMILARITY: CONTAINS 6 IQ DOMAINS.
 CC -1- SIMILARITY: CONTAINS 1 DILUTE DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: U90942; AAD00702.1; -;
 CC EMBL: Y07759; CA69035.1; -;
 CC EMBL: Y07759; CA69036.1; -;
 CC EMBL: 222957; CA68053.1; -;
 CC EMBL: S74799; AAB3211.1; -;
 CC EMBL: AF055459; AAC14188.1; -;
 CC HSSP: P08799; 1MND.
 CC MIM: 160777; -;
 CC MIM: 214450; -;
 CC InterPro: IPR002710; DIL.
 CC InterPro: IPR000048; IQ.
 CC InterPro: IPR001609; myosin_head.
 CC Pfam: PF01843; DIL; 1.
 CC Pfam: PF00612; IQ; 6.
 CC Pfam: PF00063; myosin_head; 1.
 CC PRINTS: PR00193; MYOSINHEAVY.
 CC PRODOM: PD000355; myosin_head; 1.
 CC PRODOM: PD003376; DIL; 1.
 CC SMART: SM00015; IQ; 6.
 CC SMART: SM00242; MYSC; 1.
 CC PROSITE: PSS0096; IQ; 6.
 CC MYOSIN: Repeat: ATP-binding; Calmodulin-binding; Actin-binding;
 CC Coiled coil; Phosphorylation; Alternative splicing; Polymorphism.
 CC Colled coil; 1
 CC FT DOMAIN 766 788
 CC FT DOMAIN 789 818
 CC FT DOMAIN 819 836
 CC FT DOMAIN 837 861
 CC FT DOMAIN 862 883
 CC FT DOMAIN 885 914
 CC FT DOMAIN 914 1237
 CC FT DOMAIN 1338 1445
 CC COILED COIL (POTENTIAL).
 CC COILED COIL (POTENTIAL).

FT DOMAIN 1687 1792 DILUTE.
 FT NP_BIND 163 170 ATP (POTENTIAL).
 FT DOMAIN 643 665 ACTIN-BINDING (POTENTIAL).
 FT MOD_RES 1760 1760 PHOSPHORYLATION (POTENTIAL).
 FT VARSPPLIC 1321 1347 MISSING (IN ISOFORM 2).
 FT VARSPPLIC 1413 1413 MISSING (IN ISOFORM 3).
 FT VARIANT 1246 1246 R -> C.
 FT CONFLICT 668 668 F -> L (IN REF. 1).
 FT CONFLICT 833 833 MISSING (IN REF. 4).
 FT CONFLICT 863 863 E -> G (IN REF. 1).
 FT CONFLICT 922 922 H -> R (IN REF. 1).
 FT CONFLICT 1061 1061 V -> L (IN REF. 5).
 FT CONFLICT 1089 1089 E -> Q (IN REF. 4).
 FT CONFLICT 1177 1177 D -> E (IN REF. 5).
 FT CONFLICT 1465 1477 NIPRKEDPFGML -> SYLCACCSVTVR (IN REF. 4).
 FT CONFLICT 1471 1471 K -> N (IN REF. 5).
 FT CONFLICT 1484 1484 E -> D (IN REF. 5).
 FT SEQUENCE 1855 AA; 215419 MW; 1C55AD57285FA9EC CRC64;
 SO
 Query Match 11.8%; Score 66; DB 1; Length 1855;
 Best Local Similarity 37.0%; Pred. No. 72;
 Matches 17; Conservative 5; Mismatches 24; Indels 0; Gaps 0;
 QY 19 VNGCVSTGVAMKFEONQOTIEOTIKTKNQKEISRFGSADISF 64
 DB 183 VGSASANEVKEKVLTSNPMESIGNAKTTRNDNSRFGKYEIGF 228
 RESULT 12
 TRPD_SULSO STANDARD; PRT; 345 AA.
 AC P50384;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Anthranilate phosphoribosyltransferase (EC 2.4.2.18).
 GN TRPD OR SS00890.
 OS Sulfolobus solfataricus.
 OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
 OX NCBI_TaxID=2287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=DSM 5833 / MT-4;
 RA Tutino M.L., Cubellis M., Sanna G., Marino G.;
 RT "The tryptophan operon in Sulfolobus solfataricus.";
 RT Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 35092 / DSM 1617 / P2;
 RA MEDLINE=21195237; PubMed=11298741;
 RA She Q., Singh R.K., Contafionieri F., Zivanovic Y., Allard G.,
 RA Aweyer M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
 RA De Moors A., Erasmo G., Fletcher C., Gordon P.M.K.,
 RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
 RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
 RA Charlebois R.L., Doellittle W.F., Duguet M., Gaasterland T.,
 RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
 RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
 RT Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
 RN [3]
 RP CHARACTERIZATION, AND CRYSTALLIZATION.
 RX MEDLINE=21195237; PubMed=11298741;
 RA Ivens A., Mayans O., Szadkowski H., Wilmanns M., Kirschner K.;
 RT "Purification, characterization and crystallization of thermostable
 RT anthranilate phosphoribosyltransferase from Sulfolobus solfataricus.";
 RL Eur. J. Biochem. 268:2246-2252(2001).
 CC -1- CATALYTIC ACTIVITY: Anthranilate + phosphoribosylpyrophosphate =
 CC N-5'-phosphoribosyl-anthranilate + diphosphate.
 CC -1- PATHWAY: SECOND STEP IN BIOSYNTHESIS OF TRYPTOPHAN.


```

CC -1- SUBUNIT: HOMODIMER.
CC -1- SIMILARITY: BELONGS TO THE ANTHRANILATE PHOSPHORIBOSYLTRANSFERASE
CC FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.gib.ch).
CC -----
CC
DR EMBL; Z50014; CAA90309.1; -.
DR EMBL; AE006710; AAK41173.1; -.
DR InterPro; IPR000312; Glycos_transf_3.
DR Pfam; PF00591; Glycos_transf_3; 1.
DR Pfam; PF02885; Glycos_trans_3N; 1.
DR ProDom; PD001864; Glycos_transf_3; 1.
DR TrpCophan biosynthesis; transferase; glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 345 AA; 37573 MW; D51927F4B7AFA90 CRC64;

Query Match 11.7%; Score 65.5; DB 1; Length 345;
Best Local Similarity 21.2%; Pred. No. 13;
Matches 28; Conservative 21; Mismatches 42; Indels 41; Gaps 5;

OY 8 LTISALLTALLVLTG-----CVSTGVNVMKEONOQTIEQ 40
Db 36 ILVSAIILVALNKKSGSKNEIVGFANAMELAKIDVPNAIDTAGCGGGLGVNVTSA 95
OY 41 TIIR-----GKTNNQEISSRFGSADISFMIVIKFGHTAILAPNRMOETISLI-ISFLW 94
Db 96 ILLSLVNPVAKHGNRAVSGKSGSAD-----VLEALGYNIIVPPEAKELVKNTEVFLF 149
OY 95 VKPYRP--KNLS 104
Db 150 AQYHPAKMKNVA 161

RESULT 13
SECT_CYPAPA STANDARD; PRT; 492 AA.
AC P25014;
ID 01-MAR-1992 (Rel. 21, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Preprotein translocase sect subunit.
GN SECT.
OS Cyanophora paradoxa.
OG Cyanelle.
OC Eukaryota; Glaucocystophyceae; Cyanophoraceae; Cyanophora.
OX NCBI_TaxID=2762;
RN [1]
RP SEQUENCE FROM N.A.
RP MEDLINE=91117189; PubMed=2126059;
RA Michalowski C.B., Pfanzagl B., Loeffelhardt W., Bohnert H.J.;
RT "The cyanelle S10 spc ribosomal protein gene operon from Cyanophora
paradoxa.";
RL Mol. Gen. Genet. 224:222-231(1990).
RN [2]
RP SEQUENCE FROM N.A.
RP STRAIN=LB55 / PRINGSHEIM;
RA Stirewalt V.L., Michalowski C.B., Loeffelhardt W., Bohnert H.J.,
RA Bryant D.A.;
RT "Nucleotide sequence of the cyanelle DNA from Cyanophora paradoxa.";
RL Plant Mol. Biol. Rep. 13:327-332(1995).
RN [3]
RP SEQUENCE FROM N.A.
RP STRAIN=LB55 / PRINGSHEIM;
RA Loeffelhardt W., Stirewalt V.L., Michalowski C.B., Annarella M.,
RA Farley J.Y., Schluchter W.M., Chung S., Newmann-Spallart C.,
RA Steiner J.M., Jakovitsch J., Bohnert H.J., Bryant D.A.;
```

```

RT      "The complete sequence of the cyanelle genome of Cyanophora paradoxa:
RL      the genetic complexity of a primitive plastid.";
RL      (in) Schenk H.E.A., Herrmann R., Jeon K.W., Mueller N.E.,
RL      Schweinmler W. (eds.);
RL      Eukaryotism and Symbiosis, pp.40-48, Springer-Verlag, Heidelberg
RL      (1997).
CC      -1- FUNCTION: INVOLVED IN PROTEIN EXPORT. PROBABLY INTERACTS WITH
CC      OTHER PROTEINS TO ALLOW THE TRANSLOCATION OF PROTEINS ACROSS THE
CC      CYANELLE ENDOPLASMIC RETICULUM (CER) MEMBRANES.
CC      -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. CYANELLE.
CC      -1- SIMILARITY: BELONGS TO THE SECY/SEC61-ALPHA FAMILY.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@sib-sib.ch).
-----
CC      EMBL; M30487; AAA63630.1; -.
DR      EMBL; U30821; AAA81218.1; -.
DR      Mendel; 7931; CYAPA;secy;1.
DR      InterPro: IPR002208; Secy.
DR      Pfam; PF00344; secy; 1.
DR      PRINTS; PR00303; SECYTRNLCAE.
DR      PROSITE; PS00755; SECY_1; 1.
DR      PROSITE; PS00756; SECY_2; 1.
KW      Protein transport; Transmembrane; Cyanelle; Translocation.
SQ      SEQUENCE 492 AA; 55926 MW; 8736219112BDFCD CMC64;

Query Match          11.6%; Score 64.5; DB 1; Length 492;
Best Local Similarity 41.4%; Pred. No. 24;
Matches 12; Conservative 7; Mismatches 9; Indels 1; Gaps 1;

QY      82 WOELSLIISFLMVKPYRPKNLSFLLTK 110
        | |::|::|::|::| | 1:1 |
DB      161 WAFIQSIIVISWIMIRPY-ALNMDPFLGLK 188

RESULT 14
ID Y406_MYCPN STANDARD: PRT; 157 AA.
AC Q50325;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein MG406 homolog (Cl2_orf157L).
GN MPN605 OR MP237.
OS Mycoplasma pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Mollicutes;
OC Mycoplasmataceae; Mycoplasma.
OX NCBI_TaxID=2104;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=96177562; PubMed=8604303;
RA Hilbert H., Himmelreich R., Plagens H., Herrmann R.;
RT "Sequence analysis of 56 kb from the genome of the bacterium
RT Mycoplasma pneumoniae comprising the dnaA region, the atp operon and a
RL cluster of ribosomal protein genes.";
RN Nucleic Acids Res. 24:628-639(1996).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 29342 / M129;
RX MEDLINE=97103885; PubMed=8948633;
RA Himmelreich R., Hilbert H., Plagens H., Pirkl E., Li B.-C.,
RA Herrmann R.;
RT "Complete sequence analysis of the genome of the bacterium Mycoplasma
RT pneumoniae.";
RN Nucleic Acids Res. 24:4420-4449(1996).
RL -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

```

```
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: UA3738; AAC43652.1; -.
DR EMBL: AB000024; AAC34742.1; -.
KW Hypothetical protein; Transmembrane; Complete proteome.
FT TRANSMEM 13 33 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 81 101 POTENTIAL.
FT TRANSMEM 112 132 POTENTIAL.
SQ SEQUENCE 157 AA; 18291 MW; 7D9097A405DAD3C4 CRC64;

Query Match 11.5%; Score 64; DB 1; Length 157;
Best Local Similarity 23.8%; Pred. No. 8;
Matches 30; Conservative 23; Mismatches 45; Indels 28; Gaps 6;

QY 1 MLNHHIRLTISALT-----TALLVTGCVSTGNVAKENQOITQITIKKTKNKQISNRF 56
   :|:::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 33 LTTQHLKVLVGLLPFPFVALSVGLKLDFFEFKFKH-----LSKQRYVTF 80

QY 57 GS---ADSIISMIYI-----KFGHT-ALLAPNRMOELISLISFLWYK-PYRPNLS 104
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 81 GSFFVARYLLYLPVLLAFIRPOLYFHYIATITITLFFPLKIVISFAMVPLEYHGNLLI 140

QY 105 FYLTAK 110
DB 141 SKLNK 146

RESULT 15
ID RPA1_SULSO STANDARD; PRT; 880 AA.
AC 0980R2:
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE DNA-directed RNA polymerase subunit A' (EC 2.7.7.6).
GN RPOA1 OR SSO0225.
OS Sulfolobus solfataricus.
OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX NCBI_TaxID=2287;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 35092 / DSM 1617 / P2;
RX MEDLINE=2132296; PubMed=11427726;
RA She O., Singh R.K., Contaloni F., Zivanovic Y., Allard G.,
RA Aways M.J., Chan-Welher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erauso G., Fletcher C., Gordon P.M.K.,
RA Helkamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Nt-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Senses C.W., Van der Oost J.;
RA "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
RA Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001)
RL "The complete genome of the crenarchaeon Sulfolobus solfataricus P2.";
CC -1- OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
CC SUBSTRATES.
CC -1- CATALYTIC ACTIVITY: N nucleoside triphosphate -> N diphosphate +
CC (RNA)(N).
CC -1- COFACTOR: ZINC (By similarity).
CC -1- SIMILARITY: THE COMBINED A'+A' SUBUNITS CORRESPOND TO THE A
CC SUBUNITS OF EUKARYOTIC RNA POLYMERASES I, II AND III AND TO THE
CC EUBACTERIAL BETA' SUBUNIT.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
```

```
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE006659; AAK40566.1; -.
DR InterPro: IPR000722; RNA_pol_A.
DR Pfam: PF00623; RNA_pol_A.
KW Transferase; DNA-directed RNA polymerase; Transcription; Zinc;
KM Zinc-finger; Complete proteome.
FT ZN_FING 58 101 C4-TYPE (POTENTIAL).
SQ SEQUENCE 880 AA; 99674 MW; 5C8AE20EFC71DE23 CRC64;

Query Match 11.5%; Score 64; DB 1; Length 880;
Best Local Similarity 30.0%; Pred. No. 52;
Matches 21; Conservative 11; Mismatches 22; Indels 16; Gaps 3;

QY 37 TIEQTIKTKNKQISNRFSGADSIISFMIVIKFGHTALLAPNRMOELISLISFLWYK 96
   :|:::| | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 510 TVKTTLL-----TKREAQOILGVAD-----VKIDLEPAILAPREYTGKQVISAFL 556

QY 97 PYRPNLSFY 106
DB 557 ---PKDFNFH 563

Search completed: July 30, 2002, 15:42:00
Job time: 184 sec
```

THIS PAGE BLANK (US...)

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:38:56 ; Search time 24.69 seconds
(without alignments)
777.741 Million cell updates/sec

Title: US-10-048-197-2

Perfect score: 558
Sequence: 1 MLNHHIRLTISALTALVLT.....FLWKPYPKNSLYLTAKA 111

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_19:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeal:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	78	14.0	299	16	097N22	097N22 streptococ
2	75	13.4	1195	13	09PUM2	09PUM2 xenopus lae
3	74	13.3	458	2	09LSC3	09LSC3 lactobacilli
4	73	13.1	112	16	09JW76	09JW76 neisseria m
5	72.5	13.0	138	16	P73777	P73777 synchocyst
6	71	12.7	568	10	09LXES	09LXES arabidopsis
7	69.5	12.5	328	15	09IU91	09IU91 human immun
8	69	12.4	168	15	09O1A9	09O1A9 human immun
9	69	12.4	530	5	001537	001537 caenorhabdi
10	68.5	12.3	169	2	09K2Z5	09K2Z5 streptomyce
11	68.5	12.3	380	5	000947	000947 tetrahymena
12	68.5	12.3	391	16	084574	084574 chlamydia t
13	68	12.2	182	6	077772	077772 oryctolagus
14	68	12.2	325	10	094B61	094B61 arabidopsis
15	68	12.2	360	15	P87977	P87977 human immun
16	68	12.2	430	16	09CE58	09CE58 lactococcus

17	68	12.2	1435	11	09J1Z2	09J1Z2 mus musculu
18	68	12.2	1444	11	09JLP0	09JLP0 mus musculu
19	68	12.2	1444	11	09J1Z1	09J1Z1 mus musculu
20	68	12.2	1445	11	099M81	099M81 mus musculu
21	68	12.2	1454	11	099M82	099M82 mus musculu
22	68	12.2	1455	11	099M80	099M80 mus musculu
23	67.5	12.1	210	16	09KD84	09KD84 bacillus ha
24	67	12.0	366	15	071102	071102 human immun
25	67	12.0	455	5	019553	019553 caenorhabdi
26	67	12.0	488	17	057576	057576 methanococ
27	67	12.0	799	16	09CH57	09CH57 lactococcus
28	67	12.0	819	16	031582	031582 bacillus su
29	67	12.0	873	15	090062	090062 human immun
30	66.5	11.9	310	16	097D49	097D49 clostridium
31	66.5	11.9	364	10	039162	039162 arabidopsis
32	66.5	11.9	364	10	09LX25	09LX25 arabidopsis
33	66.5	11.9	429	16	09ZM38	09ZM38 helicobacte
34	66.5	11.9	754	13	013131	013131 oncorhynch
35	66.5	11.9	754	13	013132	013132 oncorhynch
36	66.5	11.9	766	9	080245	080245 mycoplasma
37	66	11.8	396	10	064545	064545 arabidopsis
38	66	11.8	421	4	09BFX7	09BFX7 homo sapien
39	66	11.8	421	4	09H779	09H779 homo sapien
40	66	11.8	458	16	09ZDC5	09ZDC5 rickettsia
41	66	11.8	726	3	09HDP5	09HDP5 podospira a
42	65.5	11.7	103	2	046690	046690 escherichia
43	65.5	11.7	161	4	09Y3V4	09Y3V4 homo sapien
44	65.5	11.7	182	17	029802	029802 archaeoglob
45	65.5	11.7	199	10	09S0U2	09S0U2 arabidopsis

ALIGNMENTS

RESULT 1
ID 097N22 PRELIMINARY: PRT: 299 AA.
AC 097N22;
DT 01-OCT-2001 (TREMREL.18, Created)
DT 01-OCT-2001 (TREMREL.18, Last sequence update)
DE 01-DEC-2001 (TREMREL.19, Last annotation update)
DE CATION EFFLUX SYSTEM PROTEIN.
GN SP1857.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=TIGR4;
RX MEDLINE=1357209; PubMed=11463916;
RA Tettelin H., Nelson K.E., Paulsen I.T., Eisen J.A., Read T.D.,
RA Peterson S., Heidelberg J., DeBoy R.T., Haft D.H., Dodson R.J.,
RA Durkin A.S., Gwinn M., Kolonay J.F., Nelson W.C., Peterson J.D.,
RA Umayam L.A., White O., Salzberg S.L., Lewis M.R., Redune D.,
RA Holtzapple E., Khouri H., Wolf A.M., Uterback T.R., Hansen C.L.,
RA McDonald L.A., Feldblyum T.V., Anguoli S., Dickinson T., Hickey E.K.,
RA Holt I.E., Loftus B.J., Yang F., Smith H.O., Venter J.C.,
RA Dougherty B.A., Morrison D.A., Hollingshead S.K., Fraser C.M.,
RT "Complete genome sequence of a virulent isolate of Streptococcus
pneumoniae."
RL Science 293:498-506(2001).
DR EMBL: AE007477; AAK75929.1; .
DR TIGR: SP1857; .
DR InterPro: IPR002524; Cation_efflux.
DR Pfam: PF01545; Cation_efflux; 1.
KW Complete proteome.
SQ SEQUENCE 299 AA; 33528 MW; 62C23A92323E23DB CRC64;

Query Match 14.0%; Score 78; DB 16; Length 299;
Best Local Similarity 28.3%; Pred. No. 2.6;
Matches 36; Conservative 18; Mismatches 39; Indels 34; Gaps 6;

QY 10 ISALITPA-LVYTCVST--GNVAMKEONOQTIEQTI-----KQKTN 48
 DB 78 LCATLAVLVLTGSLVLTLENTKTLHPQPNVDEGILWLGIIATITNLASTLVGKSKTK 137
 QY 49 KOEISSRFGSADISF-----MIVVIRKFGHTAIIAPNRMOELSLIISFLVWKYRPNL 103
 DB 138 NESILSLHLEDTLGVAVILMAIYVLRFTDWYILDPE-----LLSLVISFILLSKALPR-- 190
 QY 104 SFYLTAK 110
 DB 191 -FWSTLK 196

RESULT 2
 Q9PUM2 PRELIMINARY; PRT; 1195 AA.
 AC Q9PUM2; PRELIMINARY; 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE RECEPTOR PROTEIN TYROSINE PHOSPHATASE RHO PRECURSOR (FRAGMENT).
 GN PPTPT.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
 CC Xenopodinae; Xenopus.
 NC NCB1_TaxID=8355;
 RX MEDLINE=20193505; PubMed=10727868;
 RA Johnson K.G., Holt C.E.;
 RT "Expression of CYP-alpha, LAR, PTP-delta, and PTP-rho in the
 RT developing xenopus visual system."
 RL Mech. Dev. 92:291-294(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Besco J.A., Johnson K.G., Frostholt A., Popesco M., Burghes A.H.M.,
 RA Holt C.E., Rother A.;
 RT "Xenopus laevis receptor protein tyrosine phosphatase rho."
 RT Submitted (Jul-1999) to the EMBL/Genbank/DBJ databases.
 CC -1 SIMILARITY: CONTAINS 1 MAM DOMAIN.
 DR EMBL: AF173857; AAD50295.1; -.
 DR HSSP: P28827; IRPM.
 DR InterPro: IPR003961; FN_III.
 DR InterPro: IPR000387; TYR_phosphatase.
 DR InterPro: IPR000242; Tyr_prot_phphatase.
 DR InterPro: IPR00041; fn3; 3.
 DR Pfam: PF00629; MAM_1.
 DR Pfam: PF00102; Y_phosphatase; 1.
 DR PRINTS: PR00020; MAMDOMAIN.
 DR PRINTS: PR00700; PRTYPHPTASE.
 DR SMART: SM00060; FN3; 3.
 DR SMART: SM00137; MAM; 1.
 DR SMART: SM00134; PTPC; 1.
 DR PROSITE: PS00740; MAM_2; 1; UNKNOWN_1.
 DR PROSITE: PS50060; MAM_2; 1.
 DR PROSITE: PS00383; TYR_PHOSPHATASE_1; 1.
 DR PROSITE: PS50056; TYR_PHOSPHATASE_2; 1.
 DR PROSITE: PS50055; TYR_PHOSPHATASE_PTP; 1.
 KW Glycoprotein; Hydroxylase; Signal.
 FT SIGNAL 1 20
 FT CHAIN 21 >1195
 FT RECEPTOR PROTEIN TYROSINE PHOSPHATASE
 FT NON_TER 1195 1195
 FT SEQUENCE 1195 AA; 132818 MW; 7153C6D0AAD3F07 CRC64;

*Query Match 13.4%; Score 75; DB 13; Length 1195;
 Best Local Similarity 30.0%; Pred. No. 26;
 Matches 27; Conservative 16; Mismatches 25; Indels 22; Gaps 4;

QY 15 TALLVTCVSTGNVAMKEONOQTIEQTIKTKNOELS-----SRFGSADISFMIY 67
 DB 228 TALMTYRVY-----NHRFSATVSVGETSORSISKRYCLVRSSGGSGVS -NVAEP 276
 QY 68 VIKFGHTAIIAPNRMOELSLIISFLVWKYRPNL 103
 DB 277 VKEPPTIAPP-----ELLAVGATYIMIKP 302

RESULT 3
 Q9L5C3 PRELIMINARY; PRT; 458 AA.
 ID Q9L5C3
 AC Q9L5C3; PRELIMINARY; 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE SURFACE LAYER PROTEIN.
 GN SLPM.
 OS Lactobacillus crispatus.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillaceae;
 CC Lactobacillus.
 CC NCB1_TaxID=47770;
 RX MEDLINE=20507814; PubMed=11053389;
 RA Sillanpaa J., Martinez B., Antikainen J., Toba T., Kalkkinen N.,
 RA Tankka S., Lountama K., Keranen J., Hook M., Westerlund-Wikstrom B.,
 RA Pouwels P.H., Korhonen T.K.;
 RT "Characterization of the collagen-binding S-layer protein CbsA of
 RT Lactobacillus crispatus."
 RL J. Bacteriol. 182:6440-6450(2000).
 DR EMBL: AF253043; AAF68971.1; -.
 SO SEQUENCE 458 AA; 48750 MW; 9132CABD6DA0AD2E CRC64;

Query Match 13.3%; Score 74; DB 2; Length 458;
 Best Local Similarity 27.2%; Pred. No. 11;
 Matches 22; Conservative 19; Mismatches 26; Indels 14; Gaps 3;

QY 1 MLNHIKRLTISALLVLTGCVS-----TGVMKKEONOQTIEQTIKTKNOEI 52
 DB 68 VLGNVKNVKNAGNISKATLSGSIATFGGKSYTANLRGTDN-----VLLNGRTADEL 122
 QY 53 SS-SRFGSADISFMIYIKFG 72
 DB 123 SNVAGSSNTTITKNVGNFG 143

RESULT 4
 Q9JW76 PRELIMINARY; PRT; 112 AA.
 ID Q9JW76
 AC Q9JW76; PRELIMINARY; 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
 DE HYPOTHETICAL PROTEIN NMA0510.
 GN NMA0510.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 CC NCB1_TaxID=65699;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.W., Davis P., Devlin K., Feltham D., Hamlin N., Holroyd S.,
 RA Jags K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.;
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491.";

Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;

Db 120 HGKPVVS---TQLLNGSLAEDDIVRTENTYTDNAKIIIVOLNETTEIN

RP SEQUENCE FROM N.A.
STRAIN=BRISTOL N2;
BC

8 LTISALITALLY--TGCVSINGNAMKEQNOQII EQ I LINSUNINXZ

Db 10 LLAATVLTAVLALVALGNITFDNNQYVHNVLAMDTTFKDDLMWRATSK-GLQ 68
QY 60 DSISFIVIKFGHTAILAPNRMOELSLIT---SPLWVK 96
Db 69 DTAVVALIV-----METALVALYIMATWLMIR 95

RESULT 11
ID 000947 PRELIMINARY; PRT: 380 AA.
AC 000947;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE TREMBSKELETAL (FRAGMENT).
OS Tetrahymena vorax.
OC Eukaryota; Alveolata; Ciliophora; Oligohymenophorea; Hymenostomatida;
OC Tetrahymenina; Tetrahymena.
OX NCBI_TaxID=5919;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CL;
RA Green M.;
RT "Tetraskeletal: A unique constitutively expressed structural protein
of Tetrahymena vorax.";
RL Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF003091; AAB61407.1; -
DR InterPro: IPR003015; HLH_Myc.
DR POSITIVE; PS00038; HELIX_LOOP_HELIX; UNKNOWN_1.
FT NON_TER 1
SQ SEQUENCE 380 AA; 41058 MW; 4F58C8DF5444E9A3 CRC64;

Query Match 12.38; Score 68.5; DB 5; Length 380;
Best Local Similarity 41.58; Pred. No. 35;
Matches 17; Conservative 7; Mismatches 16; Indels 1; Gaps 1;
QY 9 TISALTLALVTGCVSTGVNMAKEONQOTIEOTIKGKTNK 49
Db 282 TTNATMTAMATTTTIDMTNGVTKEEN-PTIVRTTEANTNK 321

RESULT 12
ID 084574 PRELIMINARY; PRT: 391 AA.
AC 084574;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE GEN. SECRETION PROTEIN F.
GN GSPF OR CT570.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UM-3/CX;
RA MEDLINE-99000809; PubMed-9784136.
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatunov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis.";
RL Science 282:754-759(1998).
DR EMBL: AE001327; AAC68172.1; -
DR InterPro: IPR001992; Bact_secr_systII.
DR Pfam: PF00482; GSPF_F.1.
KM Complete proteome.
SQ SEQUENCE 391 AA; 43557 MW; 7785601AAB540650 CRC64;

Query Match 12.38; Score 68.5; DB 16; Length 391;
Best Local Similarity 21.78; Pred. No. 36;

Matches 20; Conservative 22; Mismatches 39; Indels 11; Gaps 3;
QY 14 LTALLVTCVSTGVNMAKEONQOTIEOTIKGKTNKKEISRFSSADISPMI----- 66
Db 275 LIEALTLGCEAVSODFLKEELQEVY-QAVVNGSLSRSLSHRTWPKLVIGVALGEBG 333
QY 67 -VVIKFGHTAILAPNRMOELSLISFLWVK 97
Db 334 DLAVFAHVAQIVMEDIQRLTWTA--WCOP 363

RESULT 13
ID 077772 PRELIMINARY; PRT: 182 AA.
AC 077772;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE PUTATIVE PLASMINOGEN ACTIVATOR INHIBITOR-1 (FRAGMENT).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=LUNG;
RA Yin J., Idell S.;
RT "Cloning and Expression of Rabbit PAI-1,"
submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
RL - SIMILARITY: BELONGS TO THE SERPIN FAMILY.
DR EMBL: AF074325; AAC62632.1; -
DR HSSP: P05121; 1A7C.
DR InterPro: IPR000215; Serpin.
DR Pfam: PF00079; serpin.1.
DR SMART: SM00093; SERPIN; 1.
KM Serpin.
FT NON_TER 1
SQ SEQUENCE 182 AA; 20920 MW; B2B6AEBD92E217F CRC64;

Query Match 12.28; Score 68; DB 6; Length 182;
Best Local Similarity 25.08; Pred. No. 17;
Matches 26; Conservative 13; Mismatches 41; Indels 24; Gaps 3;

QY 7 RLTSALTLALVTGCVSTGVNMAKEONQOTIEOTIKGKTNKKEISRFSSADISPMI 66
Db 82 QLTLLLVNALYFNGQWKT-----PFSKSGTHHRRVFKSDSTISVPMVA 126
QY 67 VVIKFGHTAILAPN-RMOELSLISFLWVKPYRPKNLSFYLTA 109
Db 127 QTNKFNTTEFLTPDGHYDIEL-----PYHGETLSMRTAA 162

RESULT 14
ID 094B61 PRELIMINARY; PRT: 325 AA.
AC 094B61;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
DE UNKNOWN PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicot; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Bann J., Garlincl P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamuya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:39:21 ; Search time 30.09 Seconds

(without alignments)
409.744 Million cell updates/sec

Title: US-10-048-197-2

Perfect score: 111
Sequence: 1 MLNHHRLTISALLTALLVT.....FLWVKYPRKNSFLYATA 111Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 747574 seqs, 111073796 residues

Word size : 15

Total number of hits satisfying chosen parameters: 1

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 1000 summaries

Database : A.Geneseq_032802.*

```
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*
```

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111	100.0	111	22 AAB60640	Moraxella catarrhalis

ALIGNMENTS

RESULT 1
AAB60640
ID AAB60640 standard: Protein: 111 AA.
XX
AC AAB60640;

```
XX 02-MAY-2001 (first entry)
DT Moraxella catarrhalis strain ATCC43617 BASB122 protein.
XX
XX
DE BASB122 protein: strain ATCC43617; antigen; antibody; vaccine;
XX genetic immunisation; infection; upper respiratory tract; otitis media;
XX hearing loss; deafness; pneumonia; sinusitis; nosocomial infection;
XX invasive disease; antibacterial; auditory.
OS Moraxella catarrhalis.
XX
XX WO200109337-A2.
XX
XX 08-FEB-2001.
XX
XX 31-JUL-2000; 2000WO-EP07365.
XX
XX 30-JUL-1999; 99GB-0018034.
XX
XX 30-JUL-1999; 99GB-0018036.
XX
XX (SMIT) SMITHKLINE BEECHAM BIOLOGICALS.
XX
XX Thonard J;
XX
XX WPI: 2001-159874/16.
XX
XX N-PSDB: AAF59778.
XX
XX New BASB122 and BASB124 polypeptides and polynucleotides from Moraxella
XX catarrhalis strain ATCC 43617, useful as therapeutic agents or vaccines
XX against bacterial infections, e.g. otitis media or pneumonia
XX
XX Claim 4, Page 66; 75pp; English.
XX
XX The invention relates to the Moraxella catarrhalis strain ATCC43617
XX BASB122 and BASB124 proteins (AAB60640 and AAB60641, respectively)
XX and to DNA encoding them (AAF59778 and AAF59779, respectively). The
XX invention also relates to immunogenic fragments of the BASB122 and
XX BASB124 proteins, expression vectors and host cells comprising BASB122
XX or BASB124 nucleic acids, the recombinant production of BASB122 or
XX BASB124, vaccine compositions comprising the BASB122 or BASB124 proteins
XX or nucleic acids, an antibody against BASB122 or BASB124, therapeutic
XX compositions comprising an anti-BASB122 or anti-BASB124 antibody, and a
XX method of identifying a Moraxella catarrhalis infection via the
XX detection of BASB122 or BASB124 proteins or antibodies. The vaccine
XX compositions of the invention are useful as prophylactic or therapeutic
XX agents against Moraxella catarrhalis infections in mammals, particularly
XX humans. Moraxella catarrhalis is a Gram negative bacterium frequently
XX isolated from the human upper respiratory tract, which is responsible for
XX several pathological conditions. It is responsible for about 15% of
XX otitis media cases in children (which can lead to temporary or permanent
XX hearing loss). It also causes pneumonia in elderly people, and sinusitis,
XX nosocomial infections and, less frequently, invasive diseases. BASB122 or
XX BASB124 proteins or nucleotides may additionally be used in screening for
XX novel antibacterial compounds, and in the diagnosis and staging of
XX infections. The present sequence represents the Moraxella catarrhalis
XX strain ATCC43617 BASB122 protein.
XX
XX Sequence 111 AA:
SQ
```

Query Match 100.0%; Score 111; DB 22; Length 111;
Best Local Similarity 100.0%; Pred. No. 3.8e-111;
Matches 111; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MLNHHRLTISALLTALLVTCVSTGVNVMKEONQOTIEOTIKGTNRQEISSRFGSAD 60
DB |
DB 1 mlnhhrltsalltallvtcvstgvnmkeqngqtlleqtlkgttnkqelsrfgsad 60
QY 61 SISFMVIVIKFGHTALAPNRQELSLTISFLWVKYPRKNSFLYATA 111
DB |
DB 61 sisfmvivilfghtallapnrqelstlslflwvprknsflylata 111
```

Wed Jul 31 08:21:27 2002

Search completed: July 30, 2002, 15:43:08
Job time: 227 sec

us-10-048-197-2.oli.go.rag

Page 2

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:39:51 ; Search time 13.05 Seconds
(without alignments)
207.758 Million cell updates/sec

Title: US-10-048-197-2

Perfect score: 111

Sequence: 1 MLNHHIRLTISALLTALVT.....FLWVKPYRPKNLSFYLTAKA 111

Scoring table:

OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 231628 seqs, 2442594 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : Issued_Patents_AA:*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
---------------	-------	----------------	--------	----	-------------

No matches found

Search completed: July 30, 2002, 15:43:27
Job time: 216 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:41:46 ; Search time 16.2 seconds
(without alignments)
658.390 Million cell updates/sec

Title: US-10-048-197-2

Perfect score: 111

Sequence: 1 MLNHHIRLTISALLTALLVT.....FLWVKPYRPNLSFYLTAKA 111

Scoring table: OLIGO

Gapop 60.0 , Gapext 60.0

Searched: 28338 seqs, 96089334 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : PIR_71:*

1: pir1:*

2: pir2:*

3: pir3:*

4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match length	ID	Description
---------------	----------------	--------------	----	-------------

No matches found

Search completed: July 30, 2002, 15:43:50
Job time: 124 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: July 30, 2002, 15:43:31 ; Search time 11.9 Seconds
(without alignments)
361.166 Million cell updates/sec

Title: US-10-048-197-2

Perfect score: 111

Sequence: 1 MLNHHIRLTISALLTALVT.....FLWVKPYRPKNLSFYLTAKA 111

Scoring table:

OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 105224 seqs, 38719550 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query Score	Match	Length	ID	Description

No matches found					

Search completed: July 30, 2002, 15:47:02
Job time: 211 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 30, 2002, 15:43:11 ; Search time 24.65 Seconds
(without alignments)
779.004 Million cell updates/sec

Title: US-10-048-197-2
Perfect score: 111
Sequence: 1 MLNHHIRLTISALLTALLVT.....FLWKPYRPKNLSPLYTAKA 111

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 562222 seqs, 172994929 residues

Word size : 15

Total number of hits satisfying chosen parameters: 0

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 1000 summaries

Database : SPTREMBL_19:*

1: sp_archaea:*\n2: sp_bacteria:*\n3: sp_fungi:*\n4: sp_human:*\n5: sp_invertebrate:*\n6: sp_mammal:*\n7: sp_mhc:*\n8: sp_organelle:*\n9: sp_phage:*\n10: sp_plant:*\n11: sp_rodent:*\n12: sp_virus:*\n13: sp_vertebrate:*\n14: sp_unclassified:*\n15: sp_virus:*\n16: sp_bacteriap:*\n17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	ID	Description
No matches found						

Search completed: July 30, 2002, 15:46:44
Job time: 213 sec

THIS PAGE BLANK (USPTO)